高等学校试用教材

刘光中 编

凸分析与极值问题



高等教育出版社

高等学校试用教材

凸分析与极值问题

刘光中　编

高等教育出版社

内 容 提 要

本书是一本关于凸集和凸函数理论的入门书，包括凸集、凸函数、凸函数的微分、极值问题的最优性条件和广义凸函数等内容。

本书是经高等工科院校应用数学教材委员会评选、推荐的教材，适用于数学、应用数学等专业的本科生及研究生。

高等学校试用教材

凸分析与极值问题

刘光中 编

*

高等教育出版社出版

新华书店北京科技发行所发行

北京印刷一厂印装

*

开本850×1168 1/32 印张10.25 字数250 000

1991年5月第1版 1991年5月第1次印刷

印数0001— 1 520

ISBN 7-04-002662-7/O·1012

定价4.45元

前　言

凸分析作为数学的一个比较年青的分支，是在五十年代以后随着数学规划、最优控制理论、数理经济学等应用数学学科的兴起而发展起来的．在这方面的先驱和作出重要贡献的有W.Fenchel, J.Moreau, R.T.Rockafellar, A.Brosted 等人．现在，凸分析的理论已在很多领域发挥着越来越大的影响．例如，在最优化理论中，凸性曾经作为方法和假设．然而，最近十多年来在诸如数学规划理论这样一些分支中，凸性代表了处理具有广泛应用的最优化问题的自然的结构．利用凸性的条件就可以避免关于连续性和可微性这样一些很强的限制．

本书的目的，是为应用数学系高年级学生、研究生和应用数学工作者提供一本在极值问题中起关键作用的凸集和凸函数理论的入门书．从 1981 年起，王荫清教授和我在成都科技大学应用数学系为运筹专业的本科生、研究生讲授"凸分析"课程．经过多次教学实践，将所编讲义加以修改、整理和补充而编成目前这本书．它的主要内容包括凸集的结构，凸函数的运算、连续性，可微性，凸函数的 Fenchel 形式的对偶理论，不等式理论，约束极值问题及 Lagrange 乘子理论等．考虑到最优化理论的某些发展趋势，最后用一章讨论了广义凸函数研究中的一些进展．

阅读本书要求具备线性代数、解析几何、实分析、泛函分析初步及点集拓扑方面的知识．当然，从纯数学的观点看，本书中的一部分材料是比较初等的．这主要是考虑到为了使需要运筹学知识的经济工作者、工程技术人员及其他非数学工作者也能学懂大部分内容．

考虑到凸分析的应用和叙述的简洁性，本书的讨论只限制在 n 维欧氏空间，虽然很多结果在无穷维抽象空间中仍然是正确的，

本书不包括一般极值问题和最优控制问题的讨论，也不涉及属于凸分析范畴的不动点理论，及凸多值映射理论。

在形成本书的过程中得到王荫清教授的帮助和成都科技大学应用数学系运筹与控制教研室各位老师的支持，得到上海交通大学应用数学系胡毓达教授的热情指导，在此一并表示感谢。本书的部分内容参考了 Rockafellar R. T. 著«*Convex analysis*»及 Пшеничный Б. Н. 著«*Выпуклый анализ и экстремальные задачи*»两书．由于水平有限，书中谬误之处敬请批评指正．

刘 光 中

于成都科技大学应用数学系

1988年12月

目　录

第一章　凸　　集

§1　基本概念与记号

本书中所指的数都是实数，R 表示实数系，R^n 表示 n 维实向量空间，R^n 中的元素用 $x=(\xi_1,\cdots,\xi_n)$ 来表示，其中 ξ_i 表示 x 的第 i 个分量. 为了方便有时也将 $x\in R^n$ 看成空间中的一点. $\theta=(0,\cdots,0)$ 代表零向量，它对应坐标原点.

$x\geqslant\theta$ 表示 $\xi_i\geqslant 0, i=1,\cdots,n$.

$x\neq\theta$ 表示至少存在一个 i，$\xi_i\neq 0$.

$x=\theta$ 表示 $\xi_i=0, i=1,\cdots,n$.

设 $x=(\xi_1,\cdots,\xi_n)$，$y=(\eta_1,\cdots,\eta_n)$，则

$x+y=(\xi_1+\eta_1,\cdots,\xi_n+\eta_n)$.

$cx=(c\xi_1,\cdots,c\xi_n)$，其中 $c\in R$.

$x=y$ 表示，$\xi_i=\eta_i, i=1,\cdots,n$.

用 $\langle x,y\rangle$ 表示 x、y 的内积，定义为：

$$\langle x,y\rangle=\xi_1\eta_1+\cdots+\xi_n\eta_n. \tag{1}$$

具有这种内积的向量空间称为 n 维欧几里德(Euclid)空间，仍用 R^n 表示，简称 n 维欧氏空间.

R^n 中的向量 x 的欧氏模（或长度）定义为：

$$|x|=\langle x,x\rangle^{\frac{1}{2}}, \tag{2}$$

即 $|x|=\left(\sum_{i=1}^{n}\xi_i^2\right)^{\frac{1}{2}}$ 显然，$|x|\geqslant 0$；当且仅当 $x=\theta$ 时，$|x|=0$.

设 $x,y\in R^n$，定义 x,y 的距离 $d(x,y)$ 为：

$$d(x,y)=|x-y|. \tag{3}$$

设 $x\in R^n$，$S\subset R^n$，定义 x 和集合 S 的距离 $d(x,S)$ 是：

$$d(x,S)=\inf\{|x-y| \mid y\in S\}. \tag{4}$$

定理 1.1 对于距离 $d(x,y)$，下面的结论成立：

1) $d(x,y)\geqslant 0$；当且仅当 $x=y$ 时，$d(x,y)=0$.

2) $d(x,z)\leqslant d(x,y)+d(y,z)$.

3) $d(x,y)=d(y,x)$.

4) $d(x,y+z)=d(x-y,z)$.

5) 当 $\lambda\geqslant 0$时，$d(\lambda x,\lambda y)=\lambda d(x,y)$.

证明略.

定理 1.2 对于欧氏模，下面的结论成立：

1) $|\langle x,y\rangle|\leqslant|x||y|$. (Cauchy 不等式) (5)

2) $|x+y|\leqslant|x|+|y|$. (三角形不等式) (6)

证明 1) 如果 $y=\theta$，(5)式的结论显然成立，所以可设 $y\neq\theta$. 对于每个数 t，有

$$\langle x+ty,x+ty\rangle=\langle x,x\rangle+2t\langle x,y\rangle+t^2\langle y,y\rangle. \tag{7}$$

(7) 式右边是关于 t 的二次式，当

$$t=t_0=-\frac{\langle x,y\rangle}{\langle y,y\rangle}$$

时达到极小值.将 t_0 代入(7)得

$$0\leqslant|x+t_0y|^2=|x|^2-\frac{|\langle x,y\rangle|^2}{|y|^2},$$

即
$$|\langle x,y\rangle|^2\leqslant|x|^2|y|^2,$$

这和(5) 式是等价的.

2) $\langle x+y,x+y\rangle=\langle x,x\rangle+2\langle x,y\rangle+\langle y,y\rangle$,

由(5) 式得

$$|x+y|^2\leqslant|x|^2+2|x||y|+|y|^2=(|x|+|y|)^2, \tag{8}$$

这和(6)式是等价的. ▌

注 1)从证明过程知道(5)式中等号成立等价于 $|x+t_0y|=0$，即 $x+t_0y=\theta$. 所以当 $y\neq\theta$ 时，$|\langle x,y\rangle|=|x||y|$ 的充分必要

条件是 x 是 y 的数量倍数．如果 $\langle x, y\rangle=|x||y|$ 则 x 是 y 的非负数量倍数（反过来结论也成立）．

2）$y\neq\theta$ 时，(6) 式中等号成立的充分必要条件是 x 是 y 的非负数量倍数．

R^n 中线性无关的 n 个元素的集合 $\{b_1,\cdots,b_n\}$ 称为 R^n 的一组基．如果这个基还满足

$$<b_i,b_j>=\begin{cases}1 & i=j,\\ 0 & i\neq j,\end{cases}\quad i,j=1,\cdots,n$$

则称它是正交基．单位坐标向量

$$\begin{aligned}e_1&=(1,0,\cdots,0),\\ e_2&=(0,1,\cdots,0),\\ &\vdots\\ e_n&=(0,0,\cdots,1),\end{aligned}$$

构成 R^n 的标准正交基．显然，对每一个 $x=(\xi_1,\cdots,\xi_n)$，有

$$x=\sum_{i=1}^{n}\xi_i e_i. \tag{9}$$

例如，$x=(2,-1,3)=2e_1-e_2+3e_3$．

若用同一个符号 A 表示 $m\times n$ 实矩阵和相应的从 R^n 到 R^m 的线性变换：$x\longrightarrow Ax$，用 A^* 表示 A 的转置矩阵和相应的从 R^m 到 R^n 的伴随线性变换，则对于任意的 $x\in R^n$，$y^*\in R^m$，恒等式

$$\langle Ax,y^*\rangle=\langle x,A^*y^*\rangle \tag{10}$$

成立．（表示向量时，符号"＊"没有运算意义．为遵守矩阵乘法规则，全部向量均为列向量）．

设 $x_1,y_1\in R^n$，$x_1\neq y_1$，集合

$$\{x\mid x=(1-t)x_1+ty_1,t\in R\} \tag{11}$$

称为过 x_1 和 y_1 的直线．如果设 $z=y_1-x_1$，则该直线亦可表示为

$$\{x \mid x = x_1 + tz, t \in R\}. \tag{12}$$

连结 x_1 和 y_1 的线段定义为集合

$$\overline{x_1 y_1} = \{x \mid x = (1-t)x_1 + ty_1, t \in [0,1]\}. \tag{13}$$

如果 $t=\frac{1}{2}$,则 x 表示连结 x_1 和 y_1 的线段的中点,$t=\frac{1}{3}$, $t=\frac{2}{3}$分别对应该线段的两个三等分点(图 1).

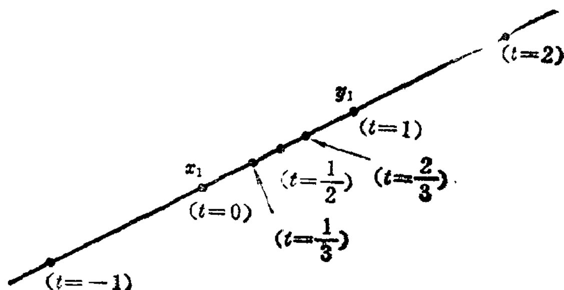


图 1

§2 R^n 中的仿射结构

这一节叙述 R^n 中的仿射理论的主要结果. 为了使叙述更加自然,首先列出 R^n 中线性结构的一些基本结论. 从这些结果以及后面将要讨论的 R^n 中的凸结构,我们就可以认识它们之间的必然联系了. 因为我们已经假定读者熟悉线性代数的有关内容,故大多数定理的证明均略去.

1. R^n 中的线性结构.

定义 2.1 设 $L \subset R^n$, 如果对于 $\forall x, y \in L, \lambda_1, \lambda_2 \in R$, 均有 $\lambda_1 x + \lambda_2 y \in L$,则称 L 是 R^n 中的子空间.

设 $x_1, \cdots, x_p$ 是 R^n 中的向量,$\lambda_1, \cdots \lambda_p \in R$,形如 $x = \lambda_1 x_1 + \cdots + \lambda_p x_p$ 的向量 x 称为 $x_1, \cdots, x_p$ 的线性组合. 所以定义 2.1 表示子空间 L 中的两个向量的任意线性组合属于 L. 事实上,下面的结果成立.

定理 2.1 设$L\subset R^n$,则L是R^n中的子空间的等价条件是L的元素的任意线性组合仍属于L.

证明略.

从线性代数知道,R^n中的子空间族的交是子空间. 所以,对于R^n中的任意集合M,存在包含M的最小子空间,这个子空间就是包含M的全体子空间的交. 我们称这个子空间是由M张成的子空间,或由M生成的子空间,或M的线性包,用$\mathrm{span}M$表示。

定理 2.2 设$M\subset R^n$,则$\mathrm{span}\,M$是M的向量的全体线性组合构成的集合.

证明略.

定义 2.2 设$\{x_1,\cdots,x_p\}$是R^n中的向量组,$\lambda_1,\cdots,\lambda_p\in R$,如果仅当$\lambda_1=\cdots=\lambda_p=0$时,有

$$\lambda_1x_1+\cdots+\lambda_px_p=\theta.$$

则称向量组$\{x_1,\cdots,x_p\}$线性无关,不是线性无关的向量组称为线性相关.

向量组的线性无关性等价于这个向量组的任何一个向量均不是向量组其他向量的线性组合. 如果某一个向量x表示成向量组$\{x_1,\cdots,x_p\}$的线性组合的形式,即$x=\lambda_1x_1+\cdots+\lambda_px_p$,则当且仅当$x_1,\cdots,x_p$线性无关时,系数$\lambda_1,\cdots,\lambda_p$可以单值地确定.

设L是R^n中的子空间,$\{x_1,\cdots,x_p\}$是R^n中的一个线性无关向量组. 如果$L=\mathrm{span}\{x_1,\cdots,x_p\}$,则称$\{x_1,\cdots,x_p\}$是$L$的一组基. L可以有无穷多个基,构成L的基的向量组中的元素并不是唯一的,但其向量的个数是相同的. L的维数$\dim L$等于L的基的元素的个数.

设$M\subset R^n$,$\dim(\mathrm{span}\,M)=p$,则存在属于M的向量组成的线性无关组$\{x_1,\cdots,x_p\}$,即$\mathrm{span}\,M$存在由M的向量组$\{x_1,\cdots,x_p\}$组成的基. 所以下面的定理成立:

理 2.3 设$M\subset R^n$，则在M中存在线性无关的向量组$\{x_1,\cdots,x_p\}$，使由形如$\sum_{i=1}^{p}\lambda_i x_i$的线性组合的全体构成的集合就是spa$M$.

这个结论表明，为了得到span$M$，只要取M中固定的线性无关向量$x_1,\cdots,x_p$的全体线性组合就可以了. 此外，span$M$中每一个向量都可以唯一地表示成$x_1,\cdots,x_p$的线性组合的形式.

定义 2.3 如果由R^n中的子空间L到R^m中的映射A保持线性组合，即

$$A\left(\sum_{i=1}^{p}\lambda_i x_i\right)=\sum_{i=1}^{p}\lambda_i A(x_i), \tag{1}$$

其中$x_i\in L,\lambda_i\in R,i=1,\cdots,p$，则称$A$是线性变换.

称由R^n中的子空间L_1到R^m中的子空间L_2上的一对一线性变换为线性同构.如果存在L_1到L_2上的线性同构，则L_1和L_2称为同构的子空间. 可以证明，当且仅当两个子空间维数相同时，这两个子空间线性同构.

2. R^n中的仿射结构

如图2，在R^2中，过原点的直线l_1是R^2中的子空间. 不过原点的直线l_2则不是子空间，但它与l_1平行，是l_1的平移. 由子空间平移而形成的这类集合将是现在研究的对象.

定义 2.4 设$M\subset R^n$. 如果对于$\forall x, y\in M, \lambda\in R$，均有$(1-\lambda)x+\lambda y\in M$，则称$M$是$R^n$中的仿射集.

有时也称仿射集为“仿射流形”、“仿射簇”、“线性流形”等.

从定义2.4可以看出，通过仿射集M中任意两点的直线仍然包含在M中.

例 1 空集$\varnothing$和空间R^n都是仿射集，R^n中的点，直线和平面都是仿射集.

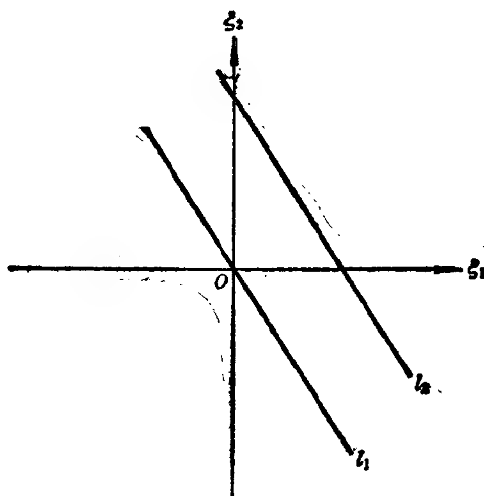


图 2

设 $x_1, \cdots, x_p$ 是 R^n 中的向量，$\lambda_1, \cdots, \lambda_p$ 是 R 中满足 $\sum_{i=1}^{p} \lambda_i = 1$ 的元素，称形如 $x = \lambda_1 x_1 + \cdots + \lambda_p x_p$ 的向量 x 是 $x_1, \cdots, x_p$ 的仿射组合．所以定义 2.4 表示，仿射集 M 中两个向量的任意仿射组合仍然属于 M．事实上，下面的结果成立．

定理 2.4 设 $M \subset R^n$，则 M 是 R^n 中的仿射集的等价条件是 M 的元素的任意仿射组合仍然属于 M．

证明略．

容易证明，R^n 中的仿射集族的交仍是 R^n 中的仿射集．所以，对于 R^n 中的任意集合 M，存在着包含 M 的最小仿射集，这个仿射集就是包含 M 的全体仿射集的交。我们称这个仿射集是由 M 张成的仿射集，或由 M 生成的仿射集，或 M 的仿射包，用 aff$M$ 表示．

例 2 如图 3，M 是 R^3 中的线段，H_1、H_2 是 R^3 中的平面，

M 在 H_1、H_2 的交线上，aff$M$ 就是 H_1、H_2 的交线。

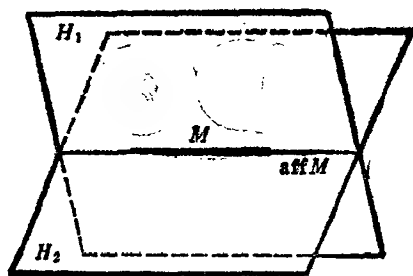


图 3

定理 2.5 设 $M \subset R^n$，则 aff M 是M的向量的全体仿射组合构成的集合.

证明留给读者.

定义 2.5 设$\{x_1, \cdots, x_p\}$是 R^n 中的向量组，$\lambda_1, \cdots, \lambda_p$ 是 R 中满足 $\lambda_1 + \cdots + \lambda_p = 0$ 的元素，如果仅当 $\lambda_1 = \cdots = \lambda_p = 0$ 时，有

$$\lambda_1 x_1 - \cdots + \lambda_p x_p = \theta,$$

则称向量组$\{x_1, \cdots x_p\}$仿射无关；不是仿射无关的向量组称为仿射相关的.

向量组的仿射无关性等价于这个向量组的任一向量都不是其他向量的仿射组合. 如果某一个向量x表示成向量组$\{x_1, \cdots, x_p\}$的仿射组合的形式，即 $x = \lambda_1 x_1 + \cdots + \lambda_p x_p$，$\lambda_1 + \cdots + \lambda_p = 1$，$\lambda_i \in R$，$i = 1, \cdots, p$，则当且仅当 $x_1, \cdots, x_p$ 仿射无关时，系数 $\lambda_1, \cdots, \lambda_p$ 可以单值地确定.

定理 2.6 设向量组$\{x_1, \cdots, x_p\} \subset R^n$，则$\{x_1, \cdots, x_p\}$仿射无关的等价条件是每一组由 $p-1$ 个向量构成的向量组

$$\{x_1 - x_i, \cdots, x_{i-1} - x_i, x_{i+1} - x_i, \cdots, x_p - x_i\},$$
$$i = 1, \cdots, p$$

线性无关.

证明留给读者.

从定理 2.6 可知,与线性无关有联系的所有事实都可以应用于仿射无关. 例如,R^n 中的任意 $p+1$ 个仿射无关的点的集合可以扩张成 $n+1$ 个仿射无关的点的集合;等等.

定理 2.7 R^n 中多于 $n+1$ 个不同向量所组成的向量组必仿射相关.

证明留给读者.

设M是 R^n 中的仿射集,$\{x_1,\cdots,x_p\}$是M中仿射无关的向量组. 如果 $M=\text{aff}\{x_1,\cdots,x_p\}$,则称$\{x_1,\cdots,x_p\}$是$M$的仿射基. 相似地定义$M$的维数 dim M等于M的仿射基的元素的个数减 1. 所以,当且仅当M的仿射基的元素个数为 p 时, $\dim M=p-1$. 当且仅当 $p=\dim M+1$ 时,M的 p 个仿射无关的向量 $x_1,\cdots,x_p$ 组成它的仿射基.

当 $M=\varnothing$时,令$\dim M=-1$. 一个点构成的仿射集是零维仿射集. 直线是一维仿射集,因为直线上不同的两个点 x_1、x_2 的集合仿射无关,故 $\text{aff}\{x_1,x_2\}$的维数等于 1, 而 $\text{aff}\{x_1,x_2\}$ 正是过 x_1、x_2 的直线.

设 $M\subset R^n$,$\dim(\text{aff}M)=p-1$,则存在属于M的向量组成的仿射无关组$\{x_1,\cdots,x_p\}$,即在$\text{aff}M$中存在由M的向量 $x_1,\cdots,x_p$ 组成的仿射基. 所以下面的结论成立:

定理 2.8 设$M\subset R^n$,则在M中存在仿射无关的向量组$\{x_1,\cdots,x_p\}$,使形如$\sum\limits_{i=1}^{p}\lambda_i x_i$的仿射组合的全体构成的集合就是 $\text{aff}M$,其中$\sum\limits_{i=1}^{p}\lambda_i=1,\lambda_i\in R,i=1,\cdots,p$.

这个结论表明,为了得到 $\text{aff}M$,只要取M中固定的仿射无关向量 $x_1,\cdots,x_p$ 的全体仿射组合就可以了. 此外, $\text{aff}M$ 中每一

个向量都可以唯一地表示成 $x_1,\cdots,x_p$ 的仿射组合的形式.

定义 2.6 如果由 R^n 中的仿射集M到 R^m 中的映射T保持仿射组合，即

$$T\left(\sum_{i=1}^{p}\lambda_i x_i\right)=\sum_{i=1}^{p}\lambda_i T(x_i), \tag{2}$$

其中 $\sum_{i=1}^{p}\lambda_i=1,\lambda_i\in R,i=1,\cdots,p$,则称$T$是仿射变换. 如果$m=1$,则仿射变换$T$也称为仿射函数.

可以证明,如果T是仿射变换,M是 R^n 中的仿射集,则$TM=\{Tx\mid x\in M\}$是 R^m 中的仿射集. 特别地，仿射变换保持仿射包，即

$$\mathrm{aff}(TM)=T(\mathrm{aff}\,M). \tag{3}$$

也可以证明,仿射变换将线段变为线段，平行线变为平行线，两平行线段的长度之比不因仿射变换而改变.

定理 2.9 仿射变换是形如 $Tx=Ax+a$ 的映射T,其中A是从 R^n 到 R^m 的线性变换,$a\in R^m$.

证明 设T是仿射变换,令 $a=T\theta,Ax=Tx-a$,容易验证 A 是从 R^n 到 R^m 的线性变换.

反之,设A是从 R^n 到 R^m 的线性变换,$Tx=Ax+a$ ，于是对于$\forall x,y\in R^n,\lambda\in R$,有

$$\begin{aligned}T[(1-\lambda)x+\lambda y]&=(1-\lambda)Ax+\lambda Ay+a\\&=(1-\lambda)Ax+(1-\lambda)a\\&\quad+\lambda Ay+\lambda a\\&=(1-\lambda)Tx+\lambda Ty.\end{aligned}$$

所以T是仿射变换. ▌

定理 2.10 设$\{x_0,x_1,\cdots,x_p\}$和$\{x_0',x_1',\cdots x_p'\}$是 R^n 中的两个仿射无关集,则存在从 R^n 到其自身的一对一仿射变换 T,使

$Tx_i = x_i', i=0,1,\cdots,p$. 如果 $p=n$,则T 是唯一的.

证明 将已知仿射无关集扩大就可以将问题归结为 $p=n$ 的情形,故不妨设 $p=n$,由线性代数的有关结果可知,存在唯一的从 R^n 到自身的一对一线性变换 A,使 $A(x_i - x_0) = x_i' - x_0'$, $i=1,\cdots,n$. 如果设 $a = x_0' - Ax_0$,则使 $Tx = Ax + a$ 的映射T即为所求的仿射变换T. ▮

推论 2.10.1 设 M_1 和 M_2 是 R^n 中维数相同的两个仿射集,则存在从 R^n 到其自身的一对一仿射变换T,使 $TM_1 = M_2$.

证明 任意 p 维仿射集可以表示成 $p+1$ 个仿射无关点的集合的仿射包, 而仿射变换保持仿射包. ▮

称由仿射集 M_1 到仿射集 M_2 上的一对一仿射变换为仿射同构. 如果存在由 M_1 到 M_2 上的仿射同构,则称 M_1 和 M_2 是仿射同构的仿射集. 由上述定理可知, 当且仅当两个仿射集维数相同时,它们仿射同构. 任何同构的仿射集也是同胚的, 即保持拓扑结构. 所有的 p 维仿射集M都与一个具体的 p 维仿射集 R^p 仿射同构. 换言之,不仅在仿射意义上,而且在拓扑意义上可以把M和 R^p 视为“等同”的集合,同时,可以将M的任意指定点和 R^p 的任意指定点视为“等同”的点.

3. 仿射集与子空间的关系

定理 2.11 R^n 中的子空间是包含原点的仿射集. 反之,包含原点的仿射集是子空间.

证明 每一个子空间包含原点,且对加法和数乘运算封闭,因而是仿射集.

反之,设M是包含原点的仿射集. 对于$\forall x \in M$, $\lambda \in R$,由定义2.4,有

$$\lambda x = (1-\lambda)\theta + \lambda x \in M,$$

即M对数乘运算封闭. 对于 $\forall x, y \in M$, 有

$$\frac{1}{2}(x+y)=\frac{1}{2}x+\left(1-\frac{1}{2}\right)y\in M,$$

故
$$x+y=2\left[\frac{1}{2}(x+y)\right]\in M.$$

即M对加法运算封闭，所以M是子空间．▌

定义 2.7 设 $M\subset R^n, a\in R^n$，集合

$$M+a=\{x-a\mid x\in M\}$$

称为M平移 a 的集合．

仿射集的平移仍然是仿射集．如果 M_1, M_2 都是仿射集，$a\in R^n$，且 $M_1=M_2+a$，则称仿射集 M_1 平行于仿射集 M_2．

定理 2.12 每一个非空仿射集M平行于唯一的子空间L，其中

$$L=M-M=\{x-y\mid x\in M, y\in M\}. \tag{4}$$

证明 首先证明M不能平行于两个不同的子空间．设有子空间 L_1 和 L_2 平行于M，则 L_1 和 L_2 彼此平行，因而存在一个 a，使 $L_2=L_1+a$．而 $\theta\in L_2$，故 $-a\in L_1$，从而$a\in L_1$．所以$L_1\supset L_1+a=L_2$，类似可证 $L_2\supset L_1$．所以 $L_1=L_2$．

再证明M平行于L．对于 $\forall y\in M$，仿射集$M-y=M+(-y)$是M的平移，且包含原点．由定理 2.11 及上面证明的，它是平行于M的唯一子空间．这个子空间与 y 的选择无关，即对于$\forall y\in M$，$L=M-y$．所以这个子空间可以表示成 $L=M-M$．▌

4．超平面

定义 2.8 R^n 中的 $n-1$ 维仿射集称为超平面．

在凸分析中，超平面是一个非常重要的概念．它是 R^3 中平面的概念在 R^n 中的推广．

下面利用正交性推导超平面的表示法．我们知道，$\langle x,y\rangle=0$ 表示 $x\perp y$．已知 R^n 的子空间L，对于 $\forall y\in L$，使$x\perp L$(即 $x\perp y$)的向量 x 的集合，称为L的正交补空间，用$L^\perp$表示．当然有

$$\dim L + \dim L^{\perp} = n,$$

且$(L^{\perp})^{\perp} = L$,如果 $x_1, \cdots, x_p$ 是 L 的基，则 $x \perp L$ 等价于 $x \perp x_1, \cdots, x \perp x_p$,特别地,$R^n$ 的 $n-1$ 维子空间是一个一维 子空间的正交补．这个子空间是由一个非零向量 b 为基张成的（唯一到一个非零数量因子）．所以 $n-1$ 维子空间是形如$\{x | x \perp b\}$的集合,其中 $b \neq \theta$,由定理 2.12,超平面是这些集合的平移．而

$$\begin{aligned}\{x | x \perp b\} + a &= \{x + a | \langle x, b\rangle = 0\} \\ &= \{y | \langle y - a, b\rangle = 0\} \\ &= \{y | \langle y, b\rangle = \alpha\},\end{aligned}$$

其中 $\alpha = \langle a, b\rangle$.只要 $\lambda \in R, \lambda \neq 0$,又有

$$\{y | \langle y, b\rangle = \alpha\} = \{y | \langle \lambda y, b\rangle = \lambda \alpha\}.$$

综上述，得

定理 2.13 已知 $\alpha \in R$,非零向量 $b \in R^n$,则集合

$$H(b, \alpha) = \{x | \langle x, b\rangle = \alpha\} \tag{5}$$

是 R^n 中的一个超平面,并且每一个超平面都可以用这种方法通过 b 和 α 表示．在准确到一个非零数量因子时，这个表示法是唯一的．

定理 2.13 中的向量 b 称为超平面 $H(b, \alpha)$的法线，而$H(b, \alpha)$的其他法线是由 b 乘以正或负数因子而得到的.每一个超平面都有"两侧"正好形象地说明了这个事实．

如果 $\alpha_1 \neq \alpha_2$,则超平面 $H(b, \alpha_1)$平行于超平面 $H(b, \alpha_2)$。如果$\alpha = 0$,则 $H(b, 0)$包含原点,因而是一个 $n-1$ 维的子空间．

例 3 在 R^4 中求包含 e_1、$e_1 + 2e_2$、$e_2 + 3e_3$、$e_3 + 4e_4$ 四点的超平面H.

解 每一个 $x \in H$ 必须满足$\langle x, b\rangle = \alpha$,其中的 b、α 正是所要求的．设 $b = (\xi_1, \xi_2, \xi_3, \xi_4)$，依次取 $x = e_1, x = e_1 + 2e_2, x = e_2 + 3e_3, x = e_3 + 4e_4$ 分别与 b 作内积得

$$\alpha = \langle e_1, b\rangle = \xi_1, \qquad \alpha = \langle e_1 + 2e_2, b\rangle = \xi_1 + 2\xi_2,$$

$$\alpha=\langle e_2+3e_3,b\rangle=\xi_2+3\xi_3,$$
$$\alpha=\langle e_3+4e_4,b\rangle=\xi_3+4\xi_4.$$
解上述关于 ξ_1,ξ_2,ξ_3,ξ_4 的线性方程组，得
$$\xi_1=\alpha,\xi_2=0,\xi_3=\frac{\alpha}{3},\xi_4=\frac{\alpha}{6}.$$
为方便计，取 $\alpha=6$，最后得
$$H=\{x=(x_1,x_2,x_3,x_4)\,|\,6x_1+2x_3+x_4=6\}.$$

R^n 中的子空间可以用 n 元齐次线性方程组的解集合表示．相应地，下面的定理说明 R^n 中的仿射集可以用 n 元线性方程组的解集合表示．

定理 2.14 设 $b\in R^m$，B 是 $m\times n$ 矩阵，则集合
$$M=\{x\in R^n\,|\,Bx=b\}\tag{6}$$
是 R^n 中的一个仿射集，且每一个仿射集均可用这种方式表示．

证明 设 $x\in M,y\in M,\lambda\in R,z=(1-\lambda)x+\lambda y$．则
$$Bz=(1-\lambda)Bx+\lambda By=(1-\lambda)b+\lambda b=b.$$
即 $z\in M$．由定义 2.4，M是仿射集．

设M是 R^n 中的非空仿射真子集，L 是平行于M 的子空间，$x_1,\cdots,x_m$ 是 $L^\perp$的基，则
$$\begin{aligned}L=(L^\perp)^\perp&=\{x\,|\,x\perp x_1,\cdots,x\perp x_m\}\\&=\{x\,|\,\langle x,x_i\rangle=0,i=1,\cdots,m\}\\&=\{x\,|\,Bx=\theta\},\end{aligned}$$
其中B是其行分别为 $x_1,\cdots,x_m$ 的 $m\times n$ 矩阵．因为 M 平行于 L，则存在一个 $a\in R^n$，使
$$M=L+a=\{x\,|\,B(x-a)=\theta\}=\{x\,|\,Bx=b\},$$
其中 $b=Ba$．

设 $M=R^n$，这时取B为 $m\times n$ 零矩阵，$b=\theta$，上面的结论显然成立．

设 $M=\varnothing$，取 B 为$m\times n$零矩阵，$b\neq\theta$，上面的结论同样成立．∎

§ 3 凸集

这一节引入凸集的概念并研究它的一系列性质．在§2已经看到，线性概念和仿射概念之间完全是相似的．在凸性概念中，相似性仍然存在，但已不再是完全的而仅仅是部分相似而已．

1．定义和基本性质

定义 3.1 设 $C\subset R^n$．如果对于 $\forall x,y\in C$，$0\leqslant\lambda\leqslant 1$，有 $(1-\lambda)x+\lambda y\in C$，则称 C 是凸集．

从定义可知，凸集的特点是如果它包含任意两个不同点 x、y，它必然也包含 x、y 之间的线段，见图4．

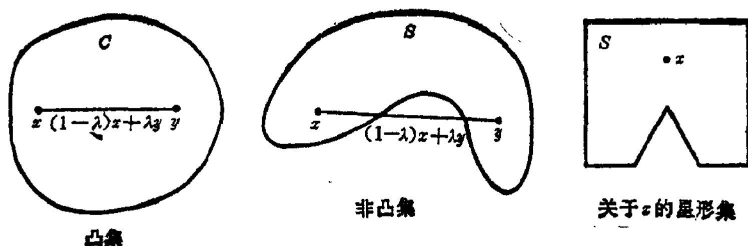


图 4

例 1 仿射集是凸集，但是凸集不一定是仿射集．例如，R^3 中的实心球、正六面体是凸集而不是仿射集．

例 2 空集 $\varnothing$，R^n，R^n 中的子空间是凸集．

例 3 超平面 $H(b,\alpha)=\{x\mid\langle x,b\rangle=\alpha\}$ 是凸集．

例 4 对于任意非零向量 $b\in R^n$ 及 $\alpha\in R$，超平面 $H(b,\alpha)=\{x\mid\langle x,b\rangle=\alpha\}$ 将 R^n 分成两个闭半空间 $\overline{K}_1(b,\alpha)=\{x\mid\langle x,b\rangle\leqslant\alpha\}$ 和 $\overline{K}_2(b,\alpha)=\{x\mid\langle x,b\rangle\geqslant\alpha\}$．而集合 $K_1(b,\alpha)=\{x\mid\langle x,b\rangle<\alpha\}$ 和 $K_2(b,\alpha)=\{x\mid\langle x,b\rangle>\alpha\}$ 称为相应的开半空间．容易验证闭、开半空间都是凸集，见图5．

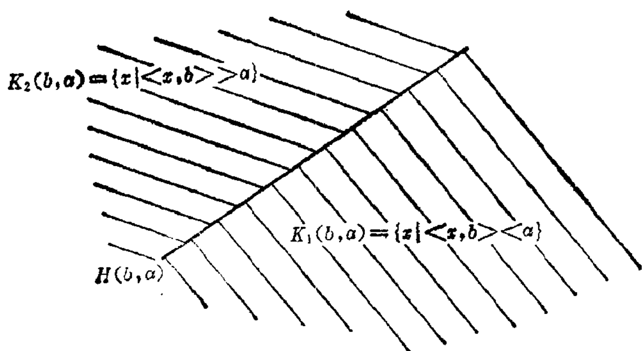


图 5

例 5 设 $x_0\in R^n,\delta>0$. 集合 $B=\{x||x-x_0|<\delta\}$ 称为以x_0为球心、δ 为半径的开球，容易验证 B 是凸集.

定义 3.2 设$C\subset R^n$, $x\in C$. 如果对于 $\forall y\in C$, $0\leqslant\lambda\leqslant1$, 有$(1-\lambda)x+\lambda y\in C$,则称$C$是关于$x$ 的星形集. 见图 4.

由定义 3.1,定义 3.2 知,R^n中的集合C是凸集的等价条件是它是关于C 的每一点的星形集. 事实上,对于任意集合C, C是关于某些点的星形集的那些点集的全体总是C 的凸子集. 下面的定义和定理表述了这个事实.

定义 3.3 设$C\subset R^n$, $\forall y\in C$,$0\leqslant\lambda\leqslant1$. 称满足$(1-\lambda)x+\lambda y\in C$ 的全体$x\in C$的集合为C 的核,记为K.

定理 3.1 设x、y、z 是 R^n中三个不同的点，$\overline{xy}$ 表示连结 x、y 的线段，$u\in\overline{xy}$,$u\neq x$, $u\neq y$. 如果$v\in\overline{zu}$, 则存在 $w\in\overline{zy}$, 使 $v\in\overline{xw}$.

证明 不失一般性,可以设 $x=\theta$. 因为 $u\in\overline{xy}$, $u\neq x$, $u\neq y$,$v\in\overline{zu}$,则存在实数 $\lambda(0<\lambda<1)$和 $\alpha(0\leqslant\alpha\leqslant1)$使 $u=\lambda y$, $v=(1-\alpha)u+\alpha z$. 容易验证

$$w=\frac{\alpha}{\alpha+\lambda(1-\alpha)}z+\frac{\lambda(1-\alpha)}{\alpha+\lambda(1-\alpha)}y\in\overline{zy},$$

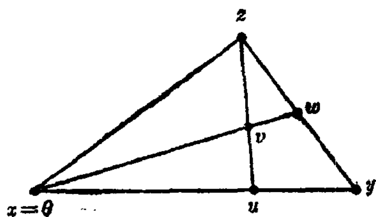


图 6

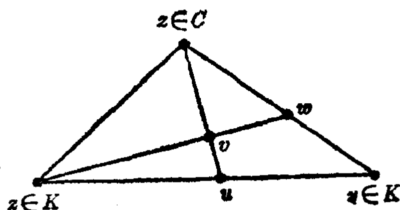


图 7

且$[\alpha+\lambda(1-\alpha)]w=v$，因而$v\in\overline{xw}$。见图6. ▌

定理 3.2 设$C\subset R^n$，则C的核K是凸集.

证明 设x、$y\in K$，$x\neq y$，$u=(1-\alpha)x+\alpha y$，$0<\alpha<1$，由定义3.3知，$u\in C$，现要证明$u\in K$.（见图7）.

设$z\in C$. 如果$z=x$或$z=y$，则$\overline{uz}\subset C$；如果$z\neq x$且$z\neq y$，令$\forall v\in\overline{uz}$. 由定理3.1，存在一点$w\in\overline{zy}$，使$v\in\overline{xw}$. 因为$y\in K$，$z\in C$，有$w\in C$. 已知$x\in K$，故$v\in C$. 因而对每一个$z\in C$，有$\overline{uz}\subseteq C$. 从而得出$u\in K$. K是凸集. ▌

定理 3.3 设I是任意指标集，$\forall i\in I$，集合C_i是R^n中的凸集，则$C=\bigcap_{i\in I}C_i$是R^n中的凸集.

证明 如果C是空集$\varnothing$或仅包含一个点，结论显然成立.

设x、$y\in C$，则对于$\forall i\in I$，x、$y\in C_i$. 有

$$(1-\lambda)x+\lambda y\in C_i,\ 0\leqslant\lambda\leqslant1,$$

故$(1-\lambda)x+\lambda y\in C$，$0\leqslant\lambda\leqslant1$. 所以$C$是凸集. ▌

推论 3.3.1 设I是可数指标集，$\forall i\in I$，集合C_i是凸集，则集合$\bigcup_{k=1}^{\infty}\bigcap_{i=k}^{\infty}C_i$是凸集.

证明留给读者.

推论 3.3.2 设I是任意指标集. $b_i\in R^n$，$\alpha_i\in R$，$i\in I$，

则集合 $C_i=\{x|\langle x,b_i\rangle\leqslant\alpha_i\}$的交是凸集，$i\in I$.

证明 因为 C_i是闭半空间，再利用定理 3.3 即可得证. ▌

如果将推论 3.3.2 中定义某些C_i的"$\leqslant$"换为"$\geqslant$"、"$<$"、"$>$"或"$=$"，结论仍然成立. 所以联立线性不等式和方程组的解集合是凸集. 这个事实在极值问题中起着重要的作用.

下面叙述凸组合的概念.

定义 3.4 设 $x_1,\cdots,x_m$是R^n中不同的点，$\lambda_i\geqslant0,i=1,\cdots,m,\sum\limits_{i=1}^{m}\lambda_i=1$. 则 $x=\sum\limits_{i=1}^{m}\lambda_ix_i$称为 $x_1,\cdots,x_m$的凸组合.

由定义 3.4 可知，凸集的定义 3.1 表示凸集是"其中任意两点的凸组合仍属于它"的集合.

例 6 设圆周 $S=\{x=(\xi,\eta)|\xi^2+\eta^2=a^2\}$，则 R^2中的圆$C=\{x=(\xi,\eta)|\xi^2+\eta^2\leqslant a^2\}$内每一点都是 S 的某两点的凸组合.

例 7 R^2中以不共线的三点 $x_0=(\xi_0,\eta_0)$，$x_1=(\xi_1,\eta_1)$，$x_2=(\xi_2,\eta_2)$为顶点的三角形上每一点都是三个顶点的凸组合. 首先，三角形每条边上的点都是两个顶点的凸组合，当然也是三个顶点的凸组合. 其次，三角形中每一点$x=(\xi,\eta)$是x_2和 x_2的对边上的点 $x_3=(\xi_3,\eta_3)$的凸组合(图 8)，而 x_3是 x_0和 x_1的凸组合，所以可以将 x 表示为三个顶点的凸组合. 即存在 $s,t\in R,0\leqslant s\leqslant1$，$0\leqslant t\leqslant1$，有

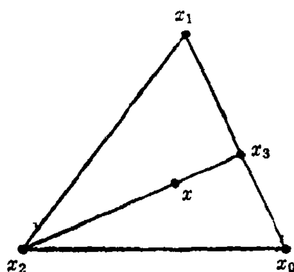


图 8

$$\begin{aligned}x&=(1-t)[(1-s)x_0+sx_1]+tx_2\\&=\lambda_0x_0+\lambda_1x_1+\lambda_2x_2,\end{aligned}$$

其中 $\lambda_0=(1-t)(1-s),\lambda_1=(1-t)s,\lambda_2=t$. 显然 $\lambda_0,\lambda_1,\lambda_2$均非

负，且 $\lambda_0+\lambda_1+\lambda_2=1$.

在定义 3.1 中，凸集是用两个点的凸组合来定义的，其实，也可以通过任意有限个点的凸组合来定义凸集．下面的定理说明了这个事实.

定理 3.4 R^n中的集合 C 是凸集的充分必要条件是它包含它的元素的所有凸组合.

证明 设 $C\subset R^n$是凸集．下面对元素个数 m 用归纳法证明 C 包含其元素的全部凸组合．$m=1$ 时结论显然成立．$m=2$ 时，由定义 3.1，结论也成立．设结论在 $m\leqslant k$ 时成立，要证明对于 $\lambda_i\geqslant 0, i=1,\cdots,k+1, \sum_{i=1}^{k+1}\lambda_i=1, x_i\in C$ 时，$x=\sum_{i=1}^{k+1}\lambda_i x_i\in C$.

不失一般性，设 $\lambda_i>0$，则 $1-\lambda_{k+1}=\sum_{i=1}^{k}\lambda_i>0$．由归纳假设

$$y=\frac{\lambda_1}{1-\lambda_{k+1}}x_1+\cdots+\frac{\lambda_k}{1-\lambda_{k+1}}x_k\in C. \tag{1}$$

因为 $\sum_{i=1}^{k}\frac{\lambda_i}{1-\lambda_{k+1}}=1$，故由定义 3.1，有

$$x=(1-\lambda_{k+1})y+\lambda_{k+1}x_{k+1}\in C. \tag{2}$$

反之，设 C 的元素的全部凸组合仍包含在 C 中．特别地，C 的任意两个元素 x_1和 x_2的凸组合也是 C 的元素，所以 C 是凸集. ▌

2. 凸包及其表示定理

根据定理 3.3，任意多个凸集的交仍是凸集，所以对于任意集合 $S\subset R^n$，存在包含 S 的最小凸集，这个凸集就是包含 S 的所有凸集的交.

定义 3.5 设 $S\subset R^n$，则包含 S 的所有凸集的交称为 S 的凸包，或由 S 张成的凸集，或由 S 生成的凸集，用 $\mathrm{co}S$ 表示.

由定理 3.3 可知，$\mathrm{co}S$ 是包含 S 的唯一最小凸集.

定理 3.5 设$S\subset R^n$,则 co$S$ 由S的点的全部凸组合构成.

证明 因为co$S$是凸集,且$S\subset$co$S$,故由定理3.4,S的点的全部凸组合也属于co$S$。若设S的点的凸组合的集合为C,则$C\subset$co$S$.

下面证明集合C是凸集. 任取$x=\lambda_1x_1+\cdots+\lambda_mx_m\in C$, $y=\mu_1y_1+\cdots+\mu_ry_r\in C$,其中$x_i,y_j\in S,i=1,\cdots,m,j=1,\cdots,r,\lambda_i\geqslant0,\mu_j\geqslant0,\sum_{i=1}^m\lambda_i=1,\sum_{j=1}^r\mu_j=1$. 当$0\leqslant\lambda\leqslant1$时,

$$(1-\lambda)x+\lambda y=(1-\lambda)\lambda_1x_1+\cdots+(1-\lambda)\lambda_mx_m+\lambda\mu_1y_1+\cdots+\lambda\mu_ry_r, \tag{3}$$

而$(1-\lambda)\lambda_1+\cdots+(1-\lambda)\lambda_m+\lambda\mu_1+\cdots+\lambda\mu_r=1$,所以$(1-\lambda)x+\lambda y$仍然是$S$的点的凸组合,故$C$是凸集.

因为$S\subset C$,co$S$是包含S的最小凸集,故co$S\subset C$. 所以co$S$的点是S的点的凸组合. ▌

推论 3.5.1 R^n中的有限子集$\{b_1,\cdots,b_m\}$的凸包由形如$\lambda_1b_1+\cdots+\lambda_mb_m$的向量构成,其中$\lambda_i\geqslant0$, $i=1,\cdots,m$, $\sum_{i=1}^m\lambda_i=1$.

证明 从$\{b_1,\cdots,b_m\}$中选出的点的每一个凸组合都可以表示成$b_1,\cdots,b_m$的凸组合. ▌

定理3.5说明S的凸包co$S$中的点是S的有限多个点的凸组合,但没有对构成这个凸组合所需要S的点的个数给出任何限制. 实际上,对于R^n中的集合S,只需要S的至多$n+1$个点的凸组合就可以表示co$S$中的点.下面的 Caratheodory 定理证明了这个事实. 这个定理在 1907 年首先被 Caratheodory 证明,是凸分析中最重要的定理之一.

定理 3.6 (Caratheodory 定理) 设$S\subset R^n$,则co$S$的点可

以表示成S的至多$n+1$个点的凸组合，即对于$\forall x\in \mathrm{co}S$，可以找到$x_1,\cdots,x_r\in S$，$\lambda_i\geqslant 0$，$i=1,\cdots,r$，$\sum_{i=1}^{r}\lambda_i=1$，$r\leqslant n+1$，使

$$x=\sum_{i=1}^{r}\lambda_i x_i. \tag{4}$$

证明 由定理3.5，只要证明$r\leqslant n+1$即可. 取形如(4)式的点. 下面证明，如果$r>n+1$，则(4)式右边非零加项可以减少.

不妨设$\lambda_i>0$，$i=1,\cdots,r$，$r>n+1$. 取$n+1$维向量$(x_i,1)$，$i=1,\cdots,r$. 因为向量组向量的个数$r>n+1$，它们线性相关. 故存在不全为零的数α_i，$i=1,\cdots,r$，使

$$\sum_{i=1}^{r}\alpha_i x_i=\theta, \tag{5}$$

$$\sum_{i=1}^{r}\alpha_i=0. \tag{6}$$

由(6)，α_i中一定存在正数. 设

$$\varepsilon_0=\min\left\{\frac{\lambda_i}{\alpha_i}\,\middle|\,\alpha_i>0,i=1,\cdots,r\right\},$$

显然，至少有一个i_0，使$\varepsilon_0=\dfrac{\lambda_{i_0}}{\alpha_{i_0}}$. 则对于$i=1,\cdots,r$，有

$$\bar{\lambda}_i=\lambda_i-\varepsilon_0\alpha_i\geqslant 0. \tag{7}$$

在$i=i_0$时，$\bar{\lambda}_{i_0}=0$.

由(5)、(6)知

$$\sum_{i=1}^{r}\bar{\lambda}_i x_i=\sum_{i=1}^{r}\lambda_i x_i-\varepsilon_0\sum_{i=1}^{r}\alpha_i x_i=x,$$

$$\sum_{i=1}^{r}\bar{\lambda}_i=\sum_{i=1}^{r}\lambda_i-\varepsilon_0\sum_{i=1}^{r}\alpha_i=1.$$

所以 x 仍然可以表示成(4)式的形式,且减少了至少一项(4)式中的非零项. 直到 $r\leqslant n+1$ 为止,均可用同样的方式继续进行. 所以定理的结论正确. ▌

在定理 3.6 的证明中,如果(4)式的 r 个向量 $x_1,\cdots,x_r$ 仿射相关,则存在不全为零的数 $\alpha_1,\cdots,\alpha_r$,使(5)、(6)两式也成立,于是 x 的形如(4)式的表达式至少可以减少一个非零项. 所以 $S\subset R^n$的凸包 $\mathrm{co}S$ 可以由 S 中的仿射无关元素的全部凸组合构成. 为了作出凸包 $\mathrm{co}S$,不一定要取 S 中点的全部凸组合,而只要作 S 中所有仿射无关点集的全部凸组合就可以了. 但是应当注意,与§2 中 $\mathrm{span}S$ 和 $\mathrm{aff}\,S$ 不同的是,这里 S 中任何一组固定的仿射无关点集都不足以构成 $\mathrm{co}S$,而是需要 S 的所有仿射无关点集的全部凸组合. 从以上说明,又可以得出定理 3.6 的一个更精确的形式.

定理 3.6′ 设 $S\subset R^n$, $\dim(\mathrm{aff}\,S)=r$,则 $\mathrm{co}S$ 的点可以表示成 S 的至多 $r+1$ 个点的凸组合.

例 8 设 x_1,x_2,x_3是 R^2中不共线的三点, x 是以 x_1, x_2, x_3 为顶点的三角形中的一点,则 x 可以表示为 x_1,x_2,x_3 的凸组合,但它不是 x_1,x_2,x_3中任意两点的凸组合.

例 8 说明在一般情况下,定理 3.6 中凸组合所需元素个数不能减少.

推论 3.6.1 设$\{C_i\subset R^n\,|\,i\in I\}$是凸集族, I 是任意指标集. $C=\mathrm{co}\left(\bigcup\limits_{i\in I}C_i\right)$.则 C 的每一点可以表示成属于不同 C_i 的至多 $n+1$ 个仿射无关点的凸组合.

证明 根据定理 3.6′,每一个 $x\in C$ 可以表示成 $\bigcup\limits_{i\in I}C_i$中 $d+1$ 个仿射无关点 $x_0,x_1,\cdots,x_d$ 的凸组合, $x=\lambda_0x_0+\lambda_1x_1+\cdots+\lambda_dx_d$. 这里 $d=\dim(\mathrm{aff}\,C)\leqslant n$,在这个凸组合中,系数为 0 的点

可以消去；如果具有非零系数的两点属于同一个 C_i，例如 x_0和x_1，令

$$y=\frac{\lambda_0}{\lambda_0+\lambda_1}x_0+\frac{\lambda_1}{\lambda_0+\lambda_1}x_1\in C_i,$$

则可将对应的两项合併为$(\lambda_0+\lambda_1)y$，且 y 和 $x_2,\cdots,x_d$仍然仿射无关．所以关于x的凸组合的表达式中 x_i 可以属于不同的凸集．▌

下面的定理是 E. Helly 在 1913 年提出并证明的．

定理 3.7(Helly 定理)　设$C_1,\cdots,C_N$是 R^n 中的凸集族，$N\geqslant n+1$．如果这个族的每一个由 $n+1$ 个凸集组成的子族的交非空，则$\bigcap_{i=1}^{N}C_i\neq\varnothing$．

证明　利用归纳法证明．$N=n+1$ 时结论显然成立．假设 $N>n+1$ 时，对于满足定理条件的$N-1$个凸集构成的凸集族结论成立．要证明满足定理条件的N个凸集组成的凸集族结论也成立．

由归纳假设，对于每一个$i=1,\cdots,N$，存在点$x^{(i)}=(\xi_1^{(i)},\cdots,\xi_n^{(i)})$，满足 $x^{(i)}\notin C_i$，但 $x^{(i)}\in C_j$，$j=1,\cdots,N$，$j\neq i$．

研究包含N个未知数 $\lambda_1,\cdots,\lambda_N$ 的 $n+1$ 个方程的线性方程组：

$$\begin{cases}\sum_{i=1}^{N}\lambda_i\xi_k^{(i)}=0,k=1,\cdots,n\\ \sum_{i=1}^{N}\lambda_i=0.\end{cases}\tag{8}$$

因为 $N>n+1$，故(8)式有非零解．用 $\lambda_{j_1},\cdots,\lambda_{j_k}$ 表示这个解中非负的λ_i(由(8)式的第$n+1$个方程知至少有一个λ_i大于零)，$\lambda_{h_1},\cdots,\lambda_{h_{N-k}}$ 表示负的λ_i．设

$$y=(y_1,\cdots,y_n),$$

其中

$$y_l=\sum_{r=1}^{k}\lambda_{j_r}\xi_l^{(j_r)}\Big/\sum_{r=1}^{k}\lambda_{j_r},l=1,\cdots,n. \qquad (9)$$

即 y 是 $x^{(j_1)},\cdots,x^{(j_k)}$的凸组合．要证明 $y\in C_i,i=1,\cdots,N$,即 y 是N个凸集的公共点．

因为 $j\neq j_r$ 时,$x^{(j_r)}\in C_j,j=1,\cdots,N,j\neq j_r$．故当 $j\neq j_1,\cdots,j_k$ 时，$x^{j_r}\in C_j,j=1,\cdots,N,j\neq j_1,\cdots,j_k$．而 y 是 $x^{(j_1)},\cdots,x^{(j_k)}$的凸组合,所以 $y\in C_j,j=1,\cdots,N,j\neq j_1,\cdots,j_k$,即 $y\in C_{h_1},\cdots,y\in C_{h_{N-k}}$．

另一方面,利用(8)式的第 $n+1$ 个方程将(9)改写成

$$y_l=\sum_{s=1}^{N-k}(-\lambda_{h_s})\xi_l^{(h_s)}\Big/\sum_{s=1}^{N-k}(-\lambda_{h_s}),l=1,\cdots,n \qquad (10)$$

于是 y 也是 $x^{(h_1)},\cdots,x^{(h_{N-k})}$的凸组合．类似可以证明 $y\in C_{j_1},\cdots,y\in C_{j_k}$．综上述,$y\in C_i,i=1,\cdots,N$．$\bigcap_{i=1}^{N}C_i\neq\varnothing$．定理结论成立．▌

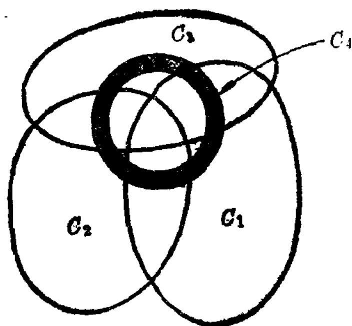


图 9

例 9 如图 9 的四个集合．其中每三个集合均有非空的交集,但四个集合的交是空集．究其原因,是 C_4不是凸集，这个例子说明定理 3.7 中所有集合均是凸集的条件是不可少的．

例 10 在 R^2 中的三角形，每两边均有公共点，但三条边却没有公共点．这说明定理 3.7 中要求每 $n+1$ 个集合的交非空这

个条件中，集合的个数 $n+1$ 是不能减少的.

注 1 我们是彼此独立地证明定理 3.6 和定理 3.7 的. 应当指出的是，Caratheodory 定理和 Helly 定理的关系是很密切的，由一个定理可以推出另一个定理. 参见文献[15].

注 2 定理 3.6 和定理 3.7 都可以推广到无界凸集，参见文献[29].

注 3 凸集的很多性质在无穷维的 Banach 空间中也是正确的，但 Helly 定理只在有限维空间中成立.

3. 单纯形

下面介绍单纯形的概念，它是一类特殊的凸集，以后还要详细地研究它.

定义 3.6 设 R^n 中的点集 $\{a_0,a_1,\cdots,a_m\}$ 仿射无关，称 $\text{co}\{a_0,a_1,\cdots,a_m\}$ 为 m 维单纯形，用 $S^m(a_0,a_1,\cdots,a_m)$ 表示. $a_0,a_1,\cdots,a_m$ 称为单纯形的顶点.

当 $m=0,1,2,3$ 时，单纯形分别是点、线段、三角形、四面体，而 R^2 中的平行四边形不是单纯形.

定义 3.7 设 $C\subset R^n$ 是凸集，$x_0\in C$. 如果 x_0 具有性质：若存在 $x,y\in C$，使 $x_0\in\overline{xy}$，则必有 $x=y=x_0$，就称 x_0 为 C 的极点.

例 11 圆的圆周上的任意一点、三角形的顶点都分别是相应凸集的极点.

定理 3.8 $a_0,a_1,\cdots,a_m$ 是 m 维单纯形 $S^m(a_0,a_1,\cdots,a_m)$ 的极点.

证明 设对于某个 k，$k\in\{0,1,\cdots,m\}$，存在 $x,y\in S^m(a_0,a_1,\cdots,a_m)$，$0<\lambda<1$，满足

$$a_k=(1-\lambda)x+\lambda y. \tag{11}$$

由定理 3.3，$x=\sum_{i=0}^{m}\alpha_i a_i$，$\alpha_i\geqslant 0$，$\sum_{i=0}^{m}\alpha_i=1$，$y=\sum_{i=0}^{m}\beta_i a_i$，$\beta_i\geqslant 0$，

$\sum_{i=0}^{m}\beta_i=1$. 故(11)式变成

$$a_k=\sum_{i=0}^{m}[(1-\lambda)\alpha_i+\lambda\beta_i]a_i. \tag{12}$$

考虑到$\sum_{i=0}^{m}[(1-\lambda)\alpha_i+\lambda\beta_i]=1$,(12)式又可写成

$$\sum_{i\neq k}[(1-\lambda)\alpha_i+\lambda\beta_i](a_i-a_k)=\theta, \tag{13}$$

或

$$\sum_{i\neq k}[(1-\lambda)\alpha_i+\lambda\beta_i](a_i-a_0)$$

$$+\left\{\sum_{i\neq k}[(1-\lambda)\alpha_i+\lambda\beta_i]\right\}(a_0-a_k)=\theta. \tag{14}$$

但 $a_1-a_0,\cdots,a_m-a_0$ 线性无关,故 $i\neq k$ 时

$$(1-\lambda)\alpha_i+\lambda\beta_i=0.$$

而 $\lambda>0,1-\lambda>0,\alpha_i\geqslant0,\beta_i\geqslant0$,故 $i\neq k$ 时

$$\alpha_i=\beta_i=0.$$

从而 $\alpha_k=\beta_k=1$. 最后得到 $x=y=a_k$. ▌

推论 3.8.1 设 $S^m(a_0,a_1,\cdots,a_m\}=S^m(b_0,b_1,\cdots,b_m)$, 则 $\{a_0,a_1,\cdots,a_m\}=\{b_0,b_1,\cdots,b_m\}$.

证明留给读者.

由定理 3.5,单纯形 $S^m(a_0,a_1,\cdots,a_m)$ 中的点 x 可以表示成其顶点 $a_0,a_1,\cdots,a_m$ 的凸组合. 其实,这个表示法也是唯一的. 事实上, 如果

$$x=\sum_{i=0}^{m}\alpha_ia_i=\sum_{i=0}^{m}\beta_ia_i. \tag{15}$$

其中 $\alpha_i,\beta_i\geqslant 0, i=0,1,\cdots,m$，$\sum_{i=0}^{m}\alpha_i=\sum_{i=0}^{m}\beta_i=1$。如果设 $\lambda_i=\alpha_i-\beta_i, i=0,1,\cdots,m$，则有$\sum_{i=0}^{m}\lambda_i a_i=\theta$。因为$\sum_{i=0}^{m}\lambda_i=0$，$\lambda_0=-(\lambda_1+\cdots+\lambda_m)$，所以 $\lambda_1(a_1-a_0)+\cdots+\lambda_m(a_m-a_0)=\theta$．但$a_1-a_0,\cdots,a_m-a_0$ 线性无关，故$\lambda_1=\cdots=\lambda_m=0$，于是 $\lambda_0=0$．最后得到 $\alpha_i=\beta_i, i=0,1,\cdots,m$．于是得到下面的定理．

定理 3.9 设$S^m(a_0,a_1,\cdots,a_m)$是 R^n 中的m维单纯形．则$S^m(a_0,a_1,\cdots,a_m)$ 中的每一点都可以唯一地表示成它的顶点的凸组合．

进一步还可以证明下面的事实：

定理 3.10 设 $C=S^m(a_0,a_1,\cdots,a_m)$，$x\in\operatorname{aff}C$．则存在 $\lambda_i\in R, i=0,1,\cdots,m$，$\sum_{i=0}^{m}\lambda_i=1$，使 $x=\sum_{i=0}^{m}\lambda_i a_i$，且这个表示式是唯一的．

证明 设$x\in\operatorname{aff}C$．由定理 2.5，存在 $x_i\in C, i=1,\cdots,k, w_i\in R$，$\sum_{i=1}^{k}w_i=1$，使 $x=\sum_{i=1}^{k}w_i x_i$．另一方面，由定理 3.9，$x_i=\sum_{j=0}^{m}\alpha_{ij}a_j, i=1,\cdots,k$，其中$\alpha_{ij}\geqslant 0, j=0,1,\cdots,m$，$\sum_{j=0}^{m}\alpha_{ij}=1$．所以有

$$x=\sum_{i=1}^{k}w_i\sum_{j=0}^{m}\alpha_{ij}a_j=\sum_{j=0}^{m}\left(\sum_{i=1}^{k}w_i\alpha_{ij}\right)a_j. \tag{16}$$

容易验证$\sum_{j=0}^{m}\sum_{i=1}^{k}w_i\alpha_{ij}=1$，如果令$\sum_{i=1}^{k}w_i\alpha_{ij}=\lambda_j$，则有

$$x=\sum_{j=0}^{m}\lambda_j a_j.$$

定理的第一部分结论得证。

再证唯一性．设

$$x=\sum_{j=0}^{m}\lambda_j a_j,\ \ \sum_{j=0}^{m}\lambda_j=1,\lambda_j\in R,j=0,1,\cdots,m, \tag{17}$$

$$x=\sum_{j=0}^{m}\mu_j a_j,\ \ \sum_{j=0}^{m}\mu_j=1,\mu_j\in R,j=0,1,\cdots,m, \tag{18}$$

则

$$\sum_{j=1}^{m}(\lambda_j-\mu_j)(a_j-a_0)=\theta. \tag{19}$$

但 $a_1-a_0,\cdots,a_m-a_0$线性无关．所以 $\lambda_j-\mu_j=0,j=1,\cdots,m$，从而有

$$\lambda_j=\mu_j,\ \ j=1,\cdots,m. \tag{20}$$

而 $\lambda_0=1-\sum_{j=1}^{m}\lambda_j=1-\sum_{j=1}^{m}\mu_j=\mu_0$．所以 x 的表示式是唯一的．▌

到现在为止，读者可能已经发现，有关子空间、仿射集的许多概念和结果与凸集中的许多概念和结果从形式上都有相似之处．但是对于凸集，基、仿射基已不再有对应的概念了．一般的凸集的点可以表示成其仿射无关的点集的凸组合，但是正如定理 3.6 后面所作的说明，这些仿射无关集的元素不是固定的．对于一类特殊的凸集——单纯形，由于它的点可以唯一地表示成其顶点的凸组合，而它的仿射包中的点也可以唯一地表示成其顶点的仿射组合，所以我们可以认为单纯形具有“凸性基”，这个“凸性基”就是它

的顶点集．下面的定理说明单纯形是唯一具有上述意义的"凸性基"的凸集．

定理 3.11 设$M=\{x_1,\cdots,x_m\}\subset R^n$，$x_1,\cdots,x_m$仿射相关．则在$M$中可以找到子集$M_1,M_2$，满足$M_1\cap M_2=\varnothing$，$M_1\cup M_2=M$，且$\mathrm{co}M_1\cap \mathrm{co}M_2\neq\varnothing$．

证明 由$x_1,\cdots,x_m$的仿射相关性，存在不全为0的实数$\lambda_1,\cdots,\lambda_m$，满足$\sum\limits_{i=1}^m\lambda_i=0$，且

$$\sum_{i=1}^m\lambda_i x_i=\theta. \tag{21}$$

设$I=\{1,\cdots,m\}$，$I_1=\{i\in I\mid\lambda_i>0\}$，$I_2=\{i\in I\mid\lambda_i\leqslant 0\}$，$M_1=\{x_i\ i\in I_1\}$，$M_2=\{x_i\mid i\in I_2\}$．显然$I_1\neq\varnothing$．

令$x=\sum\limits_{i\in I_1}\frac{\lambda_i}{\lambda}x_i$，其中$\lambda=\sum\limits_{i\in I_1}\lambda_i$．因为$\sum\limits_{i\in I_1}\frac{\lambda_i}{\lambda}=1$，$\lambda_i>0$，$i\in I_1$，所以$x\in\mathrm{co}M_1$．另一方面，由(21)又有

$$x=\sum_{i\in I_2}\frac{-\lambda_i}{\lambda}x_i, \tag{22}$$

其中$\lambda=\sum\limits_{i\in I_2}(-\lambda_i)>0$．同样，$x\in\mathrm{co}M_2$．所以$x$是$\mathrm{co}M_1$，$\mathrm{co}M_2$的公共点，即$\mathrm{co}M_1\cap\mathrm{co}M_2\neq\varnothing$．▎

推论 3.11.1 设$M=\{x_1,\cdots,x_m\}\subset R^n$，$m\geqslant n+2$．则在$M$中可以找到子集$M_1,M_2$，满足$M_1\cap M_2=\varnothing$，$M_1\cup M_2=M$，且$\mathrm{co}M_1\cap\mathrm{co}M_2\neq\varnothing$．

证明留给读者．

定义 3.8 设$C\subset R^n$，则称$\mathrm{aff}\,C$的维数是C的维数，即$\dim C=\dim(\mathrm{aff}\,C)$．

根据以上定义，不管所在的空间维数如何，圆盘总是2维的凸

集，而线段总是一维凸集.

定理 3.12 设 $A=S^m(a_0,a_1,\cdots,a_m)$，则 $\dim A=m$.

证明 设 $x\in\operatorname{aff}A$，由定理 3.10 证明中的(16)式，有

$$x-a_0=\sum_{j=0}^{m}\left(\sum_{i=1}^{k}w_i\alpha_{ij}\right)(a_j-a_0),\tag{23}$$

故

$$L=\{x-a_0\mid x\in\operatorname{aff}A\}=\operatorname{span}\{a_1-a_0,\cdots,a_m-a_0\}.$$

但 $a_1-a_0,\cdots,a_m-a_0$ 线性无关，所以 $\dim A=\dim(\operatorname{aff}A)=\dim L=m$. ∎

定理 3.13 设 $C\subset R^n$，$a_0,a_1,\cdots,a_k$ 线性无关，$a_i\in C$，$i=0,1,\cdots,k$. 则 $\dim C\geqslant k$.

证明留给读者.

定理 3.14 设 C 是 R^n 中的凸集，$\mathscr{A}$ 是包含在 C 中的全体单纯形族. 则 $\dim C=\max\{\dim A\mid A\in\mathscr{A}\}$.

证明 如果 $C=\varnothing$，结论明显成立.

设 $C\neq\varnothing$. 因为 $\dim A$ 是不超过 n 的正整数，故在 $\mathscr{A}$ 中可以找到一个单纯形 $A'=S^m(a_0,a_1,\cdots,a_m)$，使

$$\dim A'=\max\{\dim A\mid A\in\mathscr{A}\}.$$

现在证明 $C\subset\operatorname{aff}A'$. 任取 $a_{m+1}\in C$，$a_{m+1}\notin A'$. 设这样的 a_{m+1} 总存在，因为反之则有 $C=A'$，结论已经成立.

因为 C 是凸集，故 $\operatorname{co}\{a_0,a_1,\cdots,a_m,a_{m+1}\}\subset C$. 由 A' 的结构，$\operatorname{co}\{a_0,a_1,\cdots,a_m,a_{m+1}\}$ 不是单纯形，$a_1-a_0,\cdots,a_m-a_0,a_{m+1}-a_0$ 线性相关，而 $a_1-a_0,\cdots,a_m-a_0$ 线性无关，故存在不全为零的 $\lambda_1,\cdots,\lambda_m$，使

$$a_{m+1}-a_0=\sum_{i=1}^{m}\lambda_i(a_i-a_0),\tag{24}$$

$$a_{m+1}=\sum_{i=1}^{m}\lambda_i a_i+\left(1-\sum_{i=1}^{m}\lambda_i\right)a_0\in\operatorname{aff}A'.$$

由 a_{m+1} 的任意性，得

$$A' \subset C \subset \text{aff } A',$$

$$\dim A' \leqslant \dim C \leqslant \dim(\text{aff } A') = \dim A',$$

故有

$$\dim C = \dim A' = \max\{\dim A \mid A \in \mathscr{A}\}. \blacksquare$$

4. 代数运算

由定义 2.7 知，凸集 C 平移 a 是指集合

$$C + a = \{x + a \mid x \in C, a \in R^n\}. \tag{25}$$

定义 3.9 设 C 是 R^n 中的凸集，$\lambda > 0$，称

$$\lambda C = \{\lambda x \mid x \in C\} \tag{26}$$

为 C 数乘 λ.

容易证明，$C + a$ 和 λC 都是凸集.

定义 3.10 设 C_1, C_2 是 R^n 中的凸集，称

$$C_1 + C_2 = \{x_1 + x_2 \mid x_1 \in C_1, x_2 \in C_2\}$$

为 C_1, C_2 之和.

定理 3.15 设 C_1，C_2 是 R^n 中的凸集，则 $C_1 + C_2$ 也是凸集.

证明留给读者.

根据定义，集合 C 的凸性意味着

$$(1-\lambda)C + \lambda C \subset C, 0 \leqslant \lambda \leqslant 1.$$

当然，从后面的定理 3.16 可以看出，C 为凸集时等号成立.

当 $C_1, \cdots, C_m$ 是凸集，$\lambda_1, \cdots, \lambda_m \geqslant 0$ 时，由定理3.15可知 $C = \lambda_1 C_1 + \cdots + \lambda_m C_m$ 也是凸集. 为了方便，当 $\sum_{i=1}^{m} \lambda_i = 1$ 时，也称 C 为 $C_1, \cdots, C_m$ 的凸组合.

下面讨论凸集的加法和数乘的代数运算法则. 我们知道，即使没有凸性的条件，下面的结论也是成立的：

$$C_1+C_2=C_2+C_1,$$
$$(C_1+C_2)+C_3=C_1+(C_2+C_3),$$
$$\lambda_1(\lambda_2 C)=\lambda_1\lambda_2 C,\lambda_1>0,\lambda_2>0,$$
$$\lambda(C_1+C_2)=\lambda C_1+\lambda C_2,\lambda>0.$$

定理 3.16 设C是凸集，$\lambda_1\geqslant 0,\lambda_2\geqslant 0$，则

$$(\lambda_1+\lambda_2)C=\lambda_1 C+\lambda_2 C. \tag{27}$$

证明 当 $\lambda_1=\lambda_2=0$时，结论显然成立。设$\lambda_1+\lambda_2>0$，容易证明，在没有凸性限制时，关系"$\subset$"总是成立的。

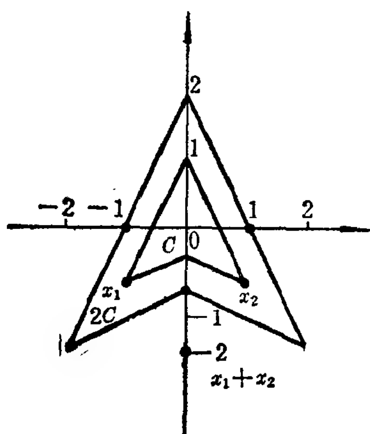


图 10

由C的凸性知，

$$\frac{\lambda_1}{\lambda_1+\lambda_2}C+\frac{\lambda_2}{\lambda_1+\lambda_2}C\subset C.$$

利用 $\lambda_1+\lambda_2$的乘法，有

$$\lambda_1 C+\lambda_2 C\subset(\lambda_1+\lambda_2)C. \tag{28}$$

故关系"$\supset$"成立。▌

容易证明，满足定理 3.16的分配律实际等价于集合的凸性。

在图10中，C不是凸集。x_1+x_2不包含在 $2C$中。这个反例说明在定理3.16中，C是凸集的条件是必不可少的。

设 C_1和 C_2 是R^n中的两个凸集，则存在一个既在 C_1中也在C_2中的最大凸集，这个凸集是$C_1\cap C_2$。也存在既包含 C_1 也包含 C_2 的最小凸集，这个凸集是 $\mathrm{co}(C_1\cup C_2)$。对于凸集族 $\{C_i|i\in I\}$，I是指标集，同样的结论也是成立的。

定理 3.17 设$\{C_i|i\in I\}$是R^n中的非空凸集族，I是指标集。令 $C=\mathrm{co}\left(\bigcup_{i\in I}C_i\right)$。则 $C=\cup\left\{\sum_{i\in I}\lambda_i C_i\right\}$，其中并的运算是全部有

限凸组合的并，即满足 $\lambda_i \geqslant 0$的系数中，只有有限个非零，且 $\sum_{i\in I}\lambda_i = 1$.

证明 根据定理3.5，C是所有凸组合 $x=\mu_1 y_1+\cdots+\mu_m y_m$ 的集合，其中 $y_1,\cdots,y_m$ 属于 $\bigcup_{i\in I} C_i$. 上面的凸组合中的向量应属于不同的 C_i，因为零向量可以不计入凸组合，而如果两个系数为正的向量属于同一个 C_i，则如推论3.6.1的证明一样，可以将相应的两项化为一项. 所以 C 是形如$\mu_1 C_{i_1}+\cdots+\mu_m C_{i_m}$ 的有限凸组合的并，其中 $i_1,\cdots,i_m$各不相同，且 $\sum_{l=1}^{m}\mu_{i_l}=1$，$\mu_{i_l}\geqslant 0$，$l=1,\cdots,m$. ▌

定义 3.11 设 A 是从 R^n到 R^m的线性变换，$C\subset R^n$，$D\subset R^m$. 称$AC=\{Ax \mid x\in C\}$是在 A 中 C的象，$A^{-1}D=\{x \mid Ax\in D\}$是在$A$中 D 的逆象($A^{-1}D$ 并不表明逆线性变换作为一个单值映射而存在).

定理 3.18 设A是从R^n到 R^m的线性变换，C，D 分别是R^n，R^m中的凸子集. 则 AC，$A^{-1}D$分别是 R^m，R^n 中的凸集.

证明留给读者.

推论 3.18.1 凸集C在子空间L 上的正交投影是一个凸集.

证明 L上的正交投影映射是线性变换，它使每一个 x 对应唯一的 y，且$(x-y)\perp L$. ▌

对于线性方程组$Ax=y$，A 是 $m\times n$ 矩阵，$x\in R^n$，$y\in R^m$. 如果 $D\subset R^m$，$y\in D$，则它的解 $x\in A^{-1}D$. 当 y 在凸集 D 上变化时，$Ax=y$ 的解集构成一个凸集$A^{-1}D$.

定义 3.12 设$C\subset R^n$，$D\subset R^m$. 称

$$C\oplus D=\{x=(y,z)\mid y\in C, z\in D\}$$

为 R^{n+m} 中 C 和 D 的直和.

定理 3.19 设C,D分别是 R^n 和 R^m 中的凸集. 则 $C\oplus D$是 R^{n+m}中的凸集.

证明留给读者.

定理 3.20 设 C_1,C_2是 R^{m+p}中的凸集. 令

$C=\{(y,z)\mid y\in R^m, z\in R^p, (y,z_1)\in C_1, (y,z_2)\in C_2, z_1+z_2=z\}$,

则 C是 R^{m+p}中的凸集.

证明 设$(y^1,z^1)\in C$,其中$(y^1,z_1^1)\in C_1$,$(y^1, z_2^1)\in C_2$, $z_1^1+z_2^1=z^1$. $(y^2,z^2)\in C$,其中$(y^2,z_1^2)\in C_1$,$(y^2,z_2^2)\in C_2$, $z_1^2+z_2^2=z^2$.

对于 $0\leqslant\lambda\leqslant 1$,由$C_1,C_2$的凸性,有

$$(1-\lambda)(y^1,z_1^1)+\lambda(y^2,z_1^2)=((1-\lambda)y^1+\lambda y^2,(1-\lambda)z_1^1+\lambda z_1^2)\in C_1,$$

$$(1-\lambda)(y^1,z_2^1)+\lambda(y^2,z_2^2)=((1-\lambda)y^1+\lambda y^2,(1-\lambda)z_2^1+\lambda z_2^2)\in C_2.$$

而

$$(1-\lambda)z^1+\lambda z^2=[(1-\lambda)z_1^1+\lambda z_1^2]+[(1-\lambda)z_2^1+\lambda z_2^2],$$

所以对同一个λ,有

$$(1-\lambda)(y^1,z^1)+\lambda(y^2,z^2)=((1-\lambda)y^1-\lambda y^2,(1-\lambda)z^1+\lambda z^2)\in C.$$

即 C 是凸集. ▌

§4 拓扑性质

1. 基本概念与性质

实分析中关于 R^n 中的开集、闭集、闭包、内点等通常的拓扑概念和性质在凸集中也是适用的.

R^n 中的欧式单位球是集合:

$$B=\{x\mid |x|\leqslant 1\}=\{x\mid d(x,\theta)\leqslant 1\}.$$

R^n 中中心在 $a\in R^n$,半径为 $\varepsilon>0$的球是:

$B(a,\varepsilon)=\{x\mid d(x,a)\leqslant\varepsilon\}=\{a+y\mid |y|\leqslant\varepsilon\}=a+\varepsilon B$.

R^n 中与集合 C 的距离不超过 ε 的集合是:

$\{x\mid \exists y\in C, d(x,y)\leqslant\varepsilon\}=\bigcup\{y+\varepsilon B\mid y\in C\}=C+\varepsilon B$.

定义 4.1 设 $C\subset R^n$,$x\in C$. 如果存在一个 $\varepsilon>0$,使 $x+\varepsilon B\subset C$,则称 x 是 C 的内点,其中 B 是单位球. C 的全体内点的集合称为 C 的内部,用 $\mathrm{int}\,C$ 表示.

易知

$$\mathrm{int}C=\{x\mid \exists\,\varepsilon>0, x+\varepsilon B\subset C\}. \tag{1}$$

定义 4.2 设 $C\subset R^n$, 如果存在一个点列 $\{x_k\}$, $x_k\in C$, $k=1,\cdots,n,\cdots$,满足 $\lim\limits_{k\to\infty}x_k=x$,则称 x 是 C 的极限点. C 的全体极限点的集合称为 C 的闭包,用 $\mathrm{cl}\,C$表示.

易知

$$\mathrm{cl}C=\bigcap\{C+\varepsilon B\mid \varepsilon>0\}. \tag{2}$$

定理 4.1 设 C 是 R^n 中的凸集,则 $\mathrm{int}C$ 和 $\mathrm{cl}\,C$ 都是凸集.

证明 如果 $\mathrm{int}\,C$ 和 $\mathrm{cl}\,C$ 是空集, 结论明显成立. 故可设 $\mathrm{int}C\neq\varnothing$, $\mathrm{cl}C\neq\varnothing$.

设 $x_1,x_2\in\mathrm{int}C$,则由定义 4.1,存在 $\varepsilon_1>0$, $\varepsilon_2>0$,使

$$x_1+\varepsilon_1 B\subset C,\ x_2+\varepsilon_2 B\subset C.$$

任取 λ, $0\leqslant\lambda\leqslant1$,由 C 的凸性,$(1-\lambda)x_1+\lambda x_2\in C$. 于是有

$$\begin{aligned}(1-\lambda)(x_1+\varepsilon_1 B)+\lambda(x_2+\varepsilon_2 B)&=(1-\lambda)x_1+\lambda x_2\\&\quad+[(1-\lambda)\varepsilon_1+\lambda\varepsilon_2]B\subset C,\end{aligned} \tag{3}$$

即$(1-\lambda)x_1+\lambda x_2\in\mathrm{int}C$. 故 $\mathrm{int}\,C$是凸集.

设 $x_1,x_2\in\mathrm{cl}C$,由定义4.2,存在点列$\{x_{1k}\}$和$\{x_{2k}\}$, $x_{ik}\in C$, $i=1,2$, $k=1,\cdots,n,\cdots$,使$\lim\limits_{k\to\infty}x_{ik}=x_i$. 任取 λ, $0\leqslant\lambda\leqslant1$, 则由 $(1-\lambda)x_{1k}+\lambda x_{2k}\in C$得

$$(1-\lambda)x_1+\lambda x_2=\lim_{k\to\infty}[(1-\lambda)x_{1k}+\lambda x_{2k}]\in\mathrm{cl}C. \tag{4}$$

故$\mathrm{cl}C$是凸集. ▌

2. 相对内部

在定理4.1的证明中的实质性部分,我们有$\mathrm{int}C\neq\varnothing$的假定. 但是,对于一般的凸集$C$,并不能保证$\mathrm{int}C\neq\varnothing$. 例如在$R^3$中,三角形是没有内点的. 但是,在由这个三角形张成的二维仿射集(仿射包)中,它确实包含有内点. 下面将要证明,总可以在某种意义下将一个非空凸集嵌入到R^n的子空间中去,使这个凸集关于这个子空间具有内点.

定理 4.2 设C是R^n中的非空凸集,则C或者有内点,或者C包含在一个维数较低的仿射集之中.

证明 设$x_0\in C$,研究形如$x-x_0$的向量,其中$x\in C$. 由线性代数的有关定理知,存在$r\leqslant n$个上述形式的线性无关向量

$$x_1-x_0,\cdots,x_r-x_0.$$

分两种情形讨论.

1) $r=n$. 这时n个向量$x_1-x_0,\cdots,x_n-x_0$线性无关,单纯形$S^n(x_0,x_1,\cdots,x_n)\subset C$. 故只要证明$S^n(x_0,x_1,\cdots,x_n)$具有内点即可. 下面证明

$$x=\lambda_0x_0+\lambda_1x_1+\cdots+\lambda_nx_n\in\mathrm{int}\,S^n(x_0,x_1,\cdots,x_n),$$

其中$\lambda_i>0,i=0,1,\cdots,n,\sum_{i=0}^{n}\lambda_i=1$.

研究关于$\lambda_i,i=1,\cdots,n$的线性方程组

$$x-x_0=\sum_{i=1}^{n}\lambda_i(x_i-x_0).$$

因为$x_1-x_0,\cdots,x_n-x_0$线性无关,这个方程组具有连续依赖于x的唯一解$\lambda_i(x),i=1,\cdots,n$. 特别取$x$的值为

$$\bar{x}=\bar{\lambda}_0x_0+\bar{\lambda}_1x_1+\cdots+\bar{\lambda}_nx_n,$$

其中 $\bar{\lambda}_i>0, i=0,1,\cdots,n, \sum_{i=0}^{n}\bar{\lambda}_i=1$. 故有

$$\lambda_i(\bar{x})=\bar{\lambda}_i>0, i=1,\cdots,n.$$

$$\lambda_0(\bar{x})=1-\sum_{i=1}^{n}\bar{\lambda}_i>0.$$

所以对 $\bar{x}$ 的某个邻域内的全体 x,有 $\lambda_i(x)>0, i=1,\cdots,n$,且

$$\lambda_0(x)=1-\sum_{i=1}^{n}\lambda_i(x)>0.$$

故对 $\bar{x}$ 的这个邻域内的全体 x,有

$$x=\sum_{i=0}^{n}\lambda_i(x)x_i\in S^n(x_0,x_1,\cdots,x_n).$$

这表示形如上述的 x 是$S^n(x_0,x_1,\cdots,x_n)$的内点. 定理的第一部分结论成立.

2) $r<n$,研究由$x_i-x_0, i=1,\cdots,r$张成的子空间 X°. 由x_i-x_0的选取可知,$C-x_0\subset X^\circ$,即$C\subset x_0+X^\circ$. 这里x_0+X°是r维的仿射集,即 C包含在一个 $r<n$ 维的 仿射 集中. 定理的第二部分结论成立. ▌

从定理4.2可以看出: i) $C-x_0$在 r 维子 空 间 X° 中 包含内点,这可以象证明 1) 的情形一样推证. ii) X°与 x_0及向量 $x_i-x_0, i=1,\cdots,r$ 的选择无关. 事实上,包含 $C-x_0$ 的任 何 子 空间应包含向量 x_i-x_0,因而也应包含X°. 所以X°是包含$C-x_0$的全体子空间的交. 如果对于某个 $x_0\in C$,子空间X°包含$C-x_0$,则它也包含由另一个点$\bar{x}_0\in C$ 形成的集合 $C-\bar{x}_0$. 事实上,

$$x-\bar{x}_0=(x-x_0)-(\bar{x}_0-x_0), \tag{5}$$

而X°是子空间,必然包含其中任意两个向量的差、于是 X° 同时包含 $C-x_0$和 $C-\bar{x}_0$,即 X°与 x_0 的选择无关。

定义 4.3 设 C 是R^n 中的凸集, $\forall x_0\in C$,包 含 $C-x_0$ 的全

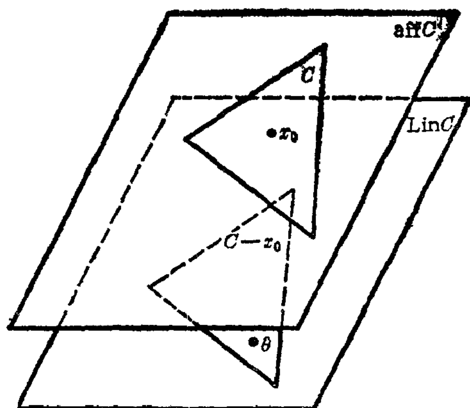


图 11

体子空间的交称为 C 的支撑子空间.用 $\mathrm{Lin}C$ 表示.

显然 $\mathrm{aff}C=x_0+\mathrm{Lin}C$. $\mathrm{aff}C$ 是 $\mathrm{Lin}C$ 平移 x_0 的集合,即 $\mathrm{aff}C$ 与 $\mathrm{Lin}C$ 平行.

在图 11 中, C 是 R^3 中的一个三角形, $\mathrm{aff}C$ 是包含 C 的一个超平面, 而 $\mathrm{Lin}C$ 则是过原点与 $\mathrm{aff}C$ 平行的一个超平面.

定义 4.4 设 C 是 R^n 中的凸集, $x\in C$. 如果存在一个 $\varepsilon>0$,使 $(x+\varepsilon B)\cap\mathrm{aff}C\subset C$ (或等价形式 $x+\mathrm{Lin}C\cap\varepsilon B\subset C$), 则称 x 是 C 的相对内点,C 的全体相对内点的集合称为 C 的相对内部, 用 $\mathrm{ri}C$ 表示.

从定义可知, 将 C 看成仿射包 $\mathrm{aff}C$ 的子集时就得出相对内部. 显然有

$$\mathrm{ri}C\subset C\subset\mathrm{cl}C.$$

集合 $\mathrm{cl}C\backslash\mathrm{ri}C$ 称为 C 的相对边界, 用 $\mathrm{rb}C$ 表示, 而 $\mathrm{rb}C$ 中的点称为相对边界点. 因为 $\mathrm{aff}C$ 是 R^n 中的闭集,所以 C 的"相对闭包"就是 C 的闭包,因而 C 的相对边界就是 C 在 $\mathrm{aff}C$ 中的边界.

如果 $\mathrm{ri}C=C$，称 C 为相对开集.

例 1 在 R^2 中研究

$$C=\{(\xi_1,0)\mid a\leqslant\xi_1\leqslant b\}.$$

显然,$\mathrm{int}\,C=\varnothing$. C 的仿射包是整个 ξ_1 轴,故

$$\mathrm{ri}\,C=\{(\xi_1,0)\mid a<\xi_1<b\}.$$

例 2 研究

$$C_1=\{x\mid\langle a,x\rangle=\alpha\},$$
$$C_2=\{x\mid\langle a,x\rangle\leqslant\alpha\}.$$

显然，$C_1\subset C_2$,故 $\mathrm{cl}C_1\subset\mathrm{cl}C_2$,$\mathrm{int}C_1\subset\mathrm{int}C_2$.而 $\mathrm{ri}C_1=C_1$, $\mathrm{ri}C_2=\{x\mid\langle a,x\rangle<\alpha\}$,所以 $\mathrm{ri}C_1\not\subset\mathrm{ri}C_2$.

取“int”运算是保持包含关系的,即 $C_1\subset C_2$ 时，一定有 $\mathrm{int}C_1\subset\mathrm{int}C_2$. 但例 2 却说明,取“ri”运算是不保持包含关系的.由此可以看出,“ri”运算绝不是“int”运算的一个无关紧要的修正而已.下面是保证 $\mathrm{ri}C_1\subset\mathrm{ri}\,C_2$成立的一个充分条件.

定理 4.3 设 $C_1\subset C_2$，$\mathrm{aff}C_1=\mathrm{aff}C_2$,则 $\mathrm{ri}C_1\subset\mathrm{ri}C_2$.

证明 设 $x\in\mathrm{ri}C_1$. 由定义 4.4,存在一个 $\varepsilon>0$,使

$$(x+\varepsilon B)\cap\mathrm{aff}C_1\subset C_1.$$

而 $\mathrm{aff}C_1=\mathrm{aff}C_2$,故

$$(x+\varepsilon B)\cap\mathrm{aff}C_2\subset C_1\subset C_2,$$

即 $x\in\mathrm{ri}C_2$,所以 $\mathrm{ri}C_1\subset\mathrm{ri}C_2$. ▌

设 C 是 R^n 中的非空凸集,可以证明,$\mathrm{ri}C=\mathrm{int}C$ 的充要条件是 $\mathrm{int}C\neq\varnothing$. 事实上,如果 $\mathrm{int}C\neq\varnothing$,则$\mathrm{aff}C=R^n$,由$\mathrm{ri}C$ 的定义可知 $\mathrm{ri}C=\mathrm{int}C$. 反过来,由下面的定理又可推出，如果 $\mathrm{ri}C=\mathrm{int}C$，则 $\mathrm{int}C\neq\varnothing$.

定理 4.4 设 C 是 R^n 中的非空凸集,则 $\mathrm{ri}C\neq\varnothing$.

证明 设 $\dim C=r$. 则在 C 中存在 $r+1$ 个仿射无关的向量 $x_0,x_1,\cdots,x_r$. 令

$$S=\mathrm{co}\{x_0,x_1,\cdots,x_r\},$$

则 S 是 r 维单纯形，$S\subset C$. 但从定理 4.2 的证明可以看出，S 具有相对于 aff S 的非空相对内部. 又因为

$$\text{aff}\, S\subset \text{aff}\, C,$$

且 $\dim(\text{aff}\, S)=r=\dim(\text{aff}\, C)$，所以有

$$\text{aff}\, S=\text{aff}\, C.$$

这表示，S 具有相对于 aff$C$ 的非空相对内部. 于是由 S 是 C 的子集可知，C 具有相对于 aff$C$ 的非空相对内部，即 $\text{ri}C\neq\varnothing$. ▌

定理 4.5 设 C 是 R^n 中的凸集，$x\in\text{ri}C$，$y\in\text{cl}C$，则 $(1-\lambda)x+\lambda y\in\text{ri}C$，$0\leqslant\lambda<1$.

证明 不妨设 $\dim C=n$，故 $\text{ri}C=\text{int}C$.

设 $0\leqslant\lambda<1$. 只要证明对于某个 $\varepsilon>0$，有

$$(1-\lambda)x+\lambda y+\varepsilon B\subset C \tag{6}$$

即可. 由 $y\in\text{cl}C$，故对于每一个 $\varepsilon>0$，$y\in C+\varepsilon B$. 于是对于每一个 $\varepsilon>0$，

$$\begin{aligned}(1-\lambda)x+\lambda y+\varepsilon B&\subset(1-\lambda)x+\lambda(C+\varepsilon B)+\varepsilon B\\&=(1-\lambda)[x+\varepsilon(1+\lambda)(1-\lambda)^{-1}B]+\lambda C.\end{aligned}$$

但 $x\in\text{int}C$，所以 ε 充分小时，

$$x+\varepsilon(1+\lambda)(1-\lambda)^{-1}B\subset C.$$

最后得

$$(1-\lambda)x+\lambda y+\varepsilon B\subset(1-\lambda)C+\lambda C=C.$$

故(6)式成立. ▌

定理 4.5 是一个很有用的工具，它在下面几个定理的证明中都起着决定性的作用.

定理 4.6 设 C 是 R^n 中的非空凸集. 则 $z\in\text{ri}C$ 的充分必要条件是对于每一个 $x\in C$，存在 $\mu>1$，使

$$(1-\mu)x+\mu z\in C.$$

证明 定理的条件表示，C 中以 z 作为端点的每一条线段可以稍微向外延长而不超出 C，见图 12. 如果 $z\in\text{ri}C$，由定义 4.4，

所述条件成立.

反之,设条件满足. 由定理 4.4, $\mathrm{ri}C\neq\varnothing$. 取 $x\in\mathrm{ri}C$. 设 $y=(1-\mu)x+\mu z\in C$, $\mu>1$. 则

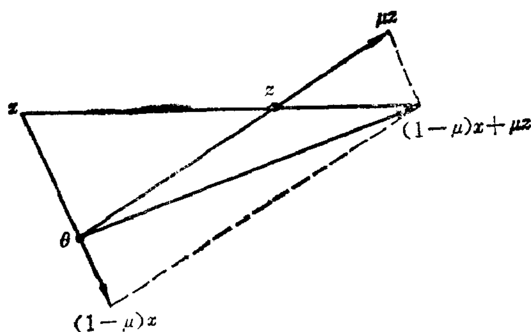


图 12

$$z=(1-\lambda)x+\lambda y,\quad 0<\lambda=\mu^{-1}<1.$$

由定理 4.5, $z\in\mathrm{ri}C$. ▌

容易证明,当 C 是凸集时, $\mathrm{ri}C$ 也是凸集.

对于 R^n 中的任意集合 C, 法则

$$\mathrm{cl}(\mathrm{cl}C)=\mathrm{cl}C,\qquad \mathrm{ri}(\mathrm{ri}C)=\mathrm{ri}C$$

总是成立的. 如果 C 是凸集,则还有下述结果.

定理 4.7 设 C 是 R^n 中的凸集, 则

1) $\mathrm{cl}C=\mathrm{cl}(\mathrm{cl}C)=\mathrm{cl}(\mathrm{ri}C)$, (7)

2) $\mathrm{ri}C=\mathrm{ri}(\mathrm{cl}C)=\mathrm{ri}(\mathrm{ri}C)$, (8)

3) $\mathrm{rb}C=\mathrm{rb}(\mathrm{cl}C)=\mathrm{rb}(\mathrm{ri}C)$, (9)

4) $\mathrm{aff}C=\mathrm{aff}(\mathrm{cl}C)=\mathrm{aff}(\mathrm{ri}C)$, (10)

5) $\dim C=\dim(\mathrm{cl}C)=\dim(\mathrm{ri}C)$. (11)

证明 如果 $C=\varnothing$,结论显然成立, 故设 $C\neq\varnothing$.

1) 因为 $\mathrm{ri}C\subset C$,故 $\mathrm{cl}(\mathrm{ri}C)\subset\mathrm{cl}C$.

设 $x_0\in\mathrm{cl}C$, $x_k\in C$, $\lambda_k\to0$, $x_k\to x_0$, $k\to\infty$. 如果 $y\in\mathrm{ri}C$, 由定理4.5, $(1-\lambda_k)x_k+\lambda_k y\in\mathrm{ri}C$. 但 $\lim\limits_{k\to\infty}(1-\lambda_k)x_k+\lambda_k y=x_0$,

所以 x_0 是 $\mathrm{ri}C$ 的极限点，即 $x_0\in\mathrm{cl}(\mathrm{ri}C)$. 所以 $\mathrm{cl}C\subset\mathrm{cl}(\mathrm{ri}C)$. (7) 式成立.

2) 因为 $\mathrm{aff}C$ 是闭集，故有 $\mathrm{aff}C=\mathrm{aff}(\mathrm{cl}C)$. 但 $C\subset\mathrm{cl}C$，由定理 4.3, $\mathrm{ri}C\subset\mathrm{ri}(\mathrm{cl}C)$.

设 $z\in\mathrm{ri}(\mathrm{cl}C)$, $\forall x\in\mathrm{ri}C$（可以设 $x\neq z$，否则结论平凡），研究过 x,z 的直线. 令 $\mu>1$，当 $\mu-1$ 充分小时直线上的点

$$y=(1-\mu)x+\mu z\in\mathrm{ri}(\mathrm{cl}C), \tag{12}$$

于是 $y\in\mathrm{cl}C$. 如果设 $\lambda=\mu^{-1}<1$，则由定理4.5,

$$z=\mu^{-1}y+(1-\mu^{-1})x=\lambda y+(1-\lambda)x\in\mathrm{ri}C. \tag{13}$$

所以 $\mathrm{ri}(\mathrm{cl}C)\subset\mathrm{ri}C$. (8)式成立.

3) 因为

$$\mathrm{rb}C=\mathrm{cl}C\backslash\mathrm{ri}C,$$
$$\mathrm{rb}(\mathrm{cl}C)=\mathrm{cl}(\mathrm{cl}C)\backslash\mathrm{ri}(\mathrm{cl}C),$$
$$\mathrm{rb}(\mathrm{ri}C)=\mathrm{cl}(\mathrm{ri}C)\backslash\mathrm{ri}(\mathrm{ri}C).$$

故由(7)、(8)式可知, (9)式成立.

4) 在2)中已经说明，由 $\mathrm{aff}C$ 是闭集而推知 $\mathrm{aff}C=\mathrm{aff}(\mathrm{cl}C)$.

将上式中的 C 换为 $\mathrm{ri}C$，利用(7)式得

$$\mathrm{aff}(\mathrm{ri}C)=\mathrm{aff}(\mathrm{cl}(\mathrm{ri}C))=\mathrm{aff}(\mathrm{cl}C)=\mathrm{aff}C.$$

故(10)式成立.

5) 由定义 3.8 及(10)式得到(11)式成立. ▌

推论4.7.1 设 C_1,C_2 是 R^n 中的凸集, $C_1\subset\mathrm{rb}C_2$. 则 $\dim C_1<\dim C_2$.

证明 设其相反, $\dim C_1=\dim C_2$. 由定义 3.8, C_1 应存在相对于 $\mathrm{aff}C_2$ 的内点. 但 $\mathrm{ri}C_2\cap C_1=\varnothing$，这样的内点不可能在 $\mathrm{cl}(\mathrm{ri}C_2)=\mathrm{cl}C_2$ 之中. 这与条件 $C_1\subset\mathrm{rb}C_2$ 矛盾，故 $\dim C_1<\dim C_2$. ▌

下面讨论在凸集族上进行运算时，相对内部的特性.

定理 4.8 设 $\{C_i\mid i\in I\}$ 是 R^n 中的凸集族, I 是任意指标集,

$\bigcap_{i\in I} \mathrm{ri}C_i \neq \varnothing$,则

$$1)\quad \mathrm{cl}\left(\bigcap_{i\in I} C_i\right)=\bigcap_{i\in I}(\mathrm{cl}C_i). \tag{14}$$

当 I 有限时，还有

$$2)\quad \bigcap_{i\in I}(\mathrm{Lin}C_i)=\mathrm{Lin}\left(\bigcap_{i\in I} C_i\right), \tag{15}$$

$$3)\quad \mathrm{ri}\left(\bigcap_{i\in I} C_i\right)=\bigcap_{i\in I}(\mathrm{ri}C_i). \tag{16}$$

证明　1) 设 $x\in\bigcap_{i\in I}(\mathrm{ri}C_i)$, $y\in\bigcap_{i\in I}(\mathrm{cl}C_i)$, 由定理4.5,对于 $0\leqslant\lambda<1$, 有

$$(1-\lambda)x+\lambda y\in\bigcap_{i\in I}(\mathrm{ri}C_i). \tag{17}$$

令 $\lambda\to1$,知 y 是 $(1-\lambda)x+\lambda y$ 的极限点,即 $y\in\mathrm{cl}\left(\bigcap_{i\in I}(\mathrm{ri}C_i)\right)$.

所以又有

$$\bigcap_{i\in I}(\mathrm{cl}C_i)\subset\mathrm{cl}\left(\bigcap_{i\in I}(\mathrm{ri}C_i)\right)\subset\mathrm{cl}\left(\bigcap_{i\in I} C_i\right)\subset\bigcap_{i\in I}(\mathrm{cl}C_i), \tag{18}$$

所以(14)成立.

在 I 有限时，不妨设 $I=\{1,2\}$, $\theta\in\mathrm{ri}C_1\cap\mathrm{ri}C_2$.

2) 因为 $\mathrm{Lin}C_1\supset C_1$, $\mathrm{Lin}C_2\supset C_2$,则

$$\mathrm{Lin}C_1\cap\mathrm{Lin}C_2\supset C_1\cap C_2.$$

故有

$$\mathrm{Lin}C_1\cap\mathrm{Lin}C_2\supset\mathrm{Lin}(C_1\cap C_2). \tag{19}$$

反之，设 $x\in\mathrm{Lin}C_1\cap\mathrm{Lin}C_2$. 因为 $\theta\in\mathrm{ri}C_1\cap\mathrm{ri}C_2$,故对于充

分小的 $\lambda>0$，$\lambda x\in C_1,\lambda x\in C_2$. 由此得

$$\lambda x\in C_1\cap C_2.$$

这表明 $\lambda x\in \mathrm{Lin}(C_1\cap C_2)$. 但 $\mathrm{Lin}(C_1\cap C_2)$ 是子空间，因而 $x\in \mathrm{Lin}(C_1\cap C_2)$，故 $\mathrm{Lin}C_1\cap \mathrm{Lin}C_2\subset \mathrm{Lin}(C_1\cap C_2)$. 所以(15)式成立.

3) 设 $x\in \mathrm{ri}C_1\cap \mathrm{ri}C_2$，则对某个 $\varepsilon>0$，有

$$x+\mathrm{Lin}C_1\cap\varepsilon B\subset C_1,\quad x+\mathrm{Lin}C_2\cap\varepsilon B\subset C_2.$$

于是

$$\begin{aligned}&(x+\mathrm{Lin}C_1\cap\varepsilon B)\cap(x+\mathrm{Lin}C_2\cap\varepsilon B)\\&=x+\mathrm{Lin}(C_1\cap C_2)\cap\varepsilon B\subset C_1\cap C_2.\end{aligned}\tag{20}$$

故 $x\in \mathrm{ri}(C_1\cap C_2)$，即 $\mathrm{ri}(C_1\cap C_2)\supset \mathrm{ri}C_1\cap \mathrm{ri}C_2$.

反之，设 $x\in \mathrm{ri}(C_1\cap C_2)$. 因为 $\theta\in \mathrm{ri}C_i, i=1,2$. 则对于 $0\leqslant\lambda<1$, $\lambda x\in \mathrm{ri}C_i, i=1,2$. 故 $\lambda x\in \mathrm{ri}C_1\cap \mathrm{ri}C_2$. 令 $\lambda\to 1$，知 x 是 $\mathrm{ri}C_1\cap \mathrm{ri}C_2$ 的极限点，即 $x\in \mathrm{cl}(\mathrm{ri}C_1\cap \mathrm{ri}C_2)$. 因而有

$$\mathrm{ri}C_1\cap \mathrm{ri}C_2\subset \mathrm{ri}(C_1\cap C_2)\subset \mathrm{cl}(\mathrm{ri}C_1\cap \mathrm{ri}C_2).\tag{21}$$

所以 $\mathrm{ri}(C_1\cap C_2)\subset \mathrm{ri}C_1\cap \mathrm{ri}C_2$. 故(16)成立. ▌

在定理 4.8 中，$\bigcap\limits_{i\in I}\mathrm{ri}C_i\neq\varnothing$ 的假设以及使(15)、(16)成立的 I 有限的假设都是必不可少的.

例 3 在 R^2 中，设 $C_1=\{(\xi_1,\xi_2)\mid\xi_1>0,\xi_2>0\}\cup\{(0,0)\}$，$C_2$ 是水平轴，即 $C_2=\{(\xi_1,0)\mid\xi_1\in R\}$. 则 $\mathrm{ri}C_1\cap \mathrm{ri}C_2=\varnothing$. $\mathrm{ri}(C_1\cap C_2)=\{(0,0)\}$，即 $\mathrm{ri}C_1\cap \mathrm{ri}C_2\neq \mathrm{ri}(C_1\cap C_2)$. 又 $\mathrm{cl}(C_1\cap C_2)=\{(0,0)\}$，而 $\mathrm{cl}C_1\cap \mathrm{cl}C_2=\{(\xi_1,0)\mid 0\leqslant\xi_1<+\infty\}$，即 $\mathrm{cl}C_1\cap \mathrm{cl}C_2\neq \mathrm{cl}(C_1\cap C_2)$.

例 4 设 $\alpha>0$. 实区间族 $[0,1+\alpha]$ 的交是区间 $[0,1]$，而 $\mathrm{ri}[0,1]=(0,1)$，区间族 $\mathrm{ri}[0,1+\alpha]$ 的交是 $(0,1]$.

推论 4.8.1 设 C 是 R^n 中的凸集，M 是至少包含 $\mathrm{ri}C$ 的一个点的仿射集，则

$$\mathrm{ri}(M\cap C)=M\cap \mathrm{ri}C,\quad \mathrm{cl}(M\cap C)=M\cap \mathrm{cl}C. \tag{22}$$

证明 因为M是仿射集，所以 $\mathrm{ri}M=M=\mathrm{cl}M$. 再利用定理4.8中的(14)、(16)即可得证. ▌

推论 4.8.2 设C_1,C_2是 R^n 中的凸集，$C_2\subset \mathrm{cl}C_1, C_2\cap \mathrm{ri}C_1\neq\varnothing$，则 $\mathrm{ri}C_2\subset \mathrm{ri}C_1$.

证明留给读者.

下面两个定理是在线性变换中闭包和相对内部的性质.

定理 4.9 设C是 R^n 中的凸集，A是从 R^n 到 R^m 的线性变换，则

$$\mathrm{ri}(AC)=A(\mathrm{ri}C), \tag{23}$$

$$\mathrm{cl}(AC)\supset A(\mathrm{cl}C). \tag{24}$$

证明 (24)式的成立与C是否为凸集无关，它只是反映了线性变换的连续性. 参考文献[38]. 下面证明(23)式的成立.

由(24)知

$$\mathrm{cl}A(\mathrm{ri}C)\supset A(\mathrm{cl}(\mathrm{ri}C))=A(\mathrm{cl}C)\supset AC\supset A(\mathrm{ri}C), \tag{25}$$

故 $\mathrm{cl}(AC)=\mathrm{cl}(A(\mathrm{ri}C))$. 从而有(见本章习题30)

$$\mathrm{ri}(AC)=\mathrm{ri}(A(\mathrm{ri}C))\subset A(\mathrm{ri}C).$$

再设 $z\in A(\mathrm{ri}C)$，要证明 $z\in \mathrm{ri}(AC)$. 设x是AC中的任意点，任意选择点 $z'\in \mathrm{ri}C, x'\in C$，使 $Az'=z, Ax'=x$. 由定理4.6，存在一个 $\mu>1$，使 $(1-\mu)x'+\mu z'\in C$. 但

$$A[(1-\mu)x'+\mu z']=(1-\mu)x+\mu z\in AC,$$

同样由定理4.6，$z\in \mathrm{ri}(AC)$. 于是 $\mathrm{ri}(AC)\supset A(\mathrm{ri}C)$，故(23)式成立. ▌

推论 4.9.1 设C是 R^n 中的凸集，$\lambda\in R$，则

$$\mathrm{ri}(\lambda C)=\lambda \mathrm{ri}C. \tag{26}$$

证明 在定理4.9中取 $A: x\to\lambda x$，即可得证. ▌

设 $C_1\subset R^m, C_2\subset R^p$ 是凸集，由定理3.19，R^{m+p}中的直和$C_1\oplus C_2$满足下面的关系：

$$\mathrm{ri}(C_1 \oplus C_2) = \mathrm{ri}C_1 \oplus \mathrm{ri}C_2, \tag{27}$$

$$\mathrm{cl}(C_1 \oplus C_2) = \mathrm{cl}C_1 \oplus \mathrm{cl}C_2. \tag{28}$$

利用(27)，(28)式可以得出下面的结论.

推论 4.9.2 设 C_1, C_2 是 R^n 中的凸集，则

$$\mathrm{ri}(C_1 + C_2) = \mathrm{ri}C_1 + \mathrm{ri}C_2, \tag{29}$$

$$\mathrm{cl}(C_1 + C_2) \supset \mathrm{cl}C_1 + \mathrm{cl}C_2. \tag{30}$$

证明 设 A 是从 R^{2n} 到 R^n 的加法线性变换，即

$$A: (x_1, x_2) \longrightarrow x_1 + x_2, x_1, x_2 \in R^n.$$

则 $A(C_1 \oplus C_2) = C_1 + C_2$. 于是利用定理 4.9 即可得到(29)，(30)式成立. ▌

定理 4.10 设 A 是从 R^n 到 R^m 的线性变换，C 是 R^m 中的凸集，$A^{-1}(\mathrm{ri}C) \neq \varnothing$. 则

$$\mathrm{ri}(A^{-1}C) = A^{-1}(\mathrm{ri}C), \tag{31}$$

$$\mathrm{cl}(A^{-1}C) = A^{-1}(\mathrm{cl}C). \tag{32}$$

证明略.

3. 凸包的拓扑性质

定理 4.11 如果 S 是 R^n 中的开集，则 $\mathrm{co}\,S$ 也是开集.

证明 不妨设 $\dim S = n$. 因为 S 是开集，故

$$S \cap \mathrm{bd}(\mathrm{co}S) = \varnothing, \tag{33}$$

其中 $\mathrm{bd}(\mathrm{co}S)$ 表示 $\mathrm{co}S$ 的边界. 但 $S \subset \mathrm{co}S$，故由(33)得到 $S \subset \mathrm{int}(\mathrm{co}S)$. 由定理4.1，$\mathrm{int}(\mathrm{co}S)$ 是凸集，于是有 $\mathrm{co}S \subset \mathrm{int}(\mathrm{co}S)$.

另一方面，$\mathrm{int}(\mathrm{co}S) \subset \mathrm{co}S$ 显然成立. 故

$$\mathrm{co}\,S = \mathrm{int}(\mathrm{co}\,S)$$

所以 $\mathrm{co}\,S$ 是开集. ▌

在§3已经说明，凸集族的交仍然是凸集. 下面将要证明，闭凸集族具有类似的性质.

定义 4.5 设 $C \subset R^n$，称以 C 为子集的全体闭凸集的交为 C 的闭凸包，用 $\mathrm{clco}\,C$ 表示.

定理 4.12 设 $C\subset R^n$，则 $\text{cl co } C=\text{cl}(\text{co}C)$.

证明 因为 $\text{co}\,C$ 是包含 C 的最小凸集，而 $\text{cl co}C$ 是包含 C 的闭凸集，故 $\text{cl co } C\supset \text{co } C$，从而 $\text{cl co } C\supset \text{cl}(\text{co } C)$.

反之，$\text{clco } C$ 是包含 C 的最小闭凸集，故 $\text{clco } C\subset \text{cl}(\text{co } C)$. 所以 $\text{cl co } C=\text{cl}(\text{co } C)$. ▌

应该注意的是，闭凸包与闭集的凸包是两回事. 下面的例子说明，闭集的凸包可能不是闭集.

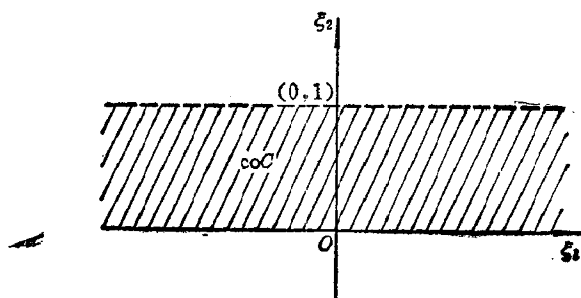


图 13

例 5 在 R^2 中，$C=\{(0,1)\}\cup\{(\xi_1,0)\mid\xi_1\in R\}$. C 是闭集，$\text{co } C=\{(\xi_1,\xi_2)\mid\xi_1\in R,0\leqslant\xi_2<1\}\cup\{(0,1)\}$. 这是一个缺上边界但包含点(0,1)的带形集，因而不是闭集. 见图 13.

定理 4.13 紧致集的凸包仍是紧致集.

证明 R^n中的紧致集是有界闭集. 设 C是紧致集，$x\in\text{co}C$. 由定理 3.6，有

$$x=\sum_{i=1}^{n+1}\lambda_i x_i,\ \ x_i\in C,\ \ \lambda_i\geqslant 0,\ \ i=1,\cdots,n+1,$$

$$\sum_{i=1}^{n+1}\lambda_i=1.$$

故$|x|\leqslant\sum_{i=1}^{n+1}\lambda_i|x_i|\leqslant K$，其中$K$是常数，满足$\forall x\in C,|x|\leqslant K$。所以co C是有界集。

再设

$$x_k=\sum_{i=1}^{n+1}\lambda_{ik}x_{ik},\ x_{ik}\in C,\ \lambda_{ik}\geqslant 0,\ i=1,\cdots,n+1,$$

$$\sum_{i=1}^{n+1}\lambda_{ik}=1.$$

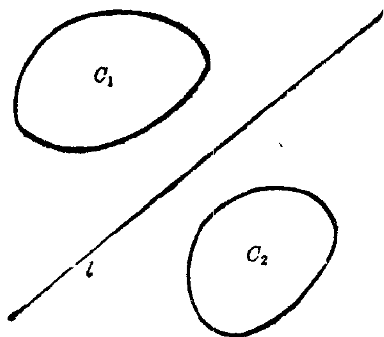


图 1.4

因为序列$\{\lambda_{ik}\}$，$\{x_{ik}\}$有界，则可从其中选出收敛子列．不失一般性，设$\lambda_{ik}\xrightarrow[k\to\infty]{}\lambda_{i0}$，$x_{ik}\xrightarrow[k\to\infty]{}x_{i0}\in C$．因为$C$是紧致集，$x_k\in$ co C．故

$$\lim_{k\to\infty}x_k=x_0=\sum_{i=1}^{n+1}\lambda_{i0}x_{i0},$$

其中$x_{i0}\in C$，$i=1,\cdots,n+1$，$\sum_{i=1}^{n+1}\lambda_{i0}=1$．这表示$x_0\in$ co C，从而co C是闭集．

§5 分离定理

在图 14 中，R^2中两个互不相交的凸集 C_1和 C_2具有一个明显的几何性质，这就是存在一条直线 l 将C_1和 C_2分开，使 C_1和 C_2分别位于 l 的两侧. 这样的几何性质在 R^n中是否有相应的推广呢？对于R^n中不相交的凸集 C_1和 C_2，是否存在一个超平面$H(a,\alpha)$，使 C_1在$H(a,\alpha)$的一侧而 C_2在$H(a,\alpha)$的另一侧呢？在一定的条件下，这个问题的答案是肯定的. 这是凸集最主要且有广泛应用的性质之一.

定义 5.1 设 C_1, C_2是 R^n 中的非空凸集，如果存在一个超平面H，使 C_1 包含在被H分成的一个闭半空间中，而 C_2 在相对的另一个闭半空间中，且 C_1, C_2 至少有一个不包含在H中，则称超平面H分离C_1和 C_2；如果 C_1包含在被H分成的一个开半空间中而 C_2包含在相对的另一个开半空间中，则称超平面H严格分离 C_1和C_2；如果存在一个 $\varepsilon>0$, $C_1+\varepsilon B$ 包含在被H分成的一个开半空间中，$C_2+\varepsilon B$包含在相对的另一个开半空间中，其中 B是欧氏单位球，则称超平面H强分离 C_1 和 C_2. 在所有情况下，H称为分离超平面.

从定义 5.1 可以看出，强分离的两个凸集一定可以严格分离和分离，严格分离的两个凸集一定可以分离. 以上关系反过来则不成立.

例 1 图 15 中，(1)的情形是H分离 C_1 和 C_2，但不能严格分离和强分离；(2)的情形H分离、严格分离 C_1 和 C_2，但不能强分离；(3)的情形H分离、严格分离、强分离 C_1和 C_2.

1. 超平面 H 分离 C_1, C_2 的条件

定理 5.1 设C_1, C_2是R^n中的非空凸集. 存在超平面H分离C_1, C_2的充分必要条件是存在一个向量 $x^*\in R^n$，满足：

1) $\inf\{\langle x,x^*\rangle\mid x\in C_1\}\geqslant\sup\{\langle x,x^*\rangle\mid x\in C_2\}$, (1)

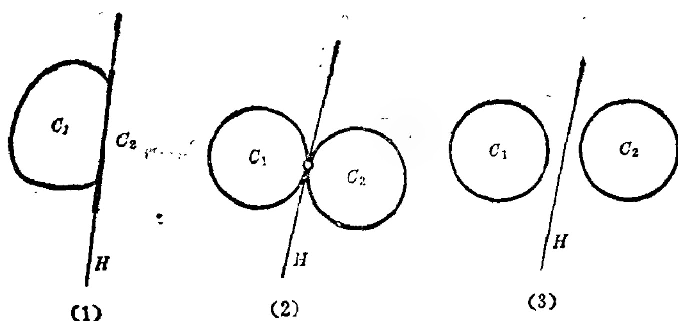


图 15

2) $\sup\{\langle x, x^*\rangle \mid x \in C_1\} > \inf\{\langle x, x^*\rangle \mid x \in C_2\}$. (2)

证明 设存在$x^* \in R^n$,使(1)、(2)式成立. 由(2)式,$x^* \neq \theta$. 由(1)式,存在 $\alpha \in R$,满足

$$\inf\{\langle x, x^*\rangle \mid x \in C_1\} \geqslant \alpha \geqslant \sup\{\langle x, x^*\rangle \mid x \in C_2\}. \tag{3}$$

由定理 2.13,$H(x^*, \alpha) = \{x \mid \langle x, x^*\rangle = \alpha\}$是一个超平面. 显然有

$$C_1 \subset \{x \mid \langle x, x^*\rangle \geqslant \alpha\}, \quad C_2 \subset \{x \mid \langle x, x^*\rangle \leqslant \alpha\},$$

即 C_1包含在闭半空间 $\overline{K}_1(x^*, \alpha) = \{x \mid \langle x, x^*\rangle \geqslant \alpha\}$之中,而 C_2包含在闭半空间 $\overline{K}_2(x^*, \alpha) = \{x \mid \langle x, x^*\rangle \leqslant \alpha\}$之中.由(2)式,$C_1, C_2$中至少有一个不包含在$H(x^*, \alpha)$之中,所以超平面 $H(x^*, \alpha)$分离 C_1和C_2.

反之,设超平面 $H(x^*, \alpha)$分离 C_1, C_2. 则由定义 5.1,有

$$C_1 \subset \overline{K}_1(x^*, \alpha) = \{x \mid \langle x, x^*\rangle \geqslant \alpha\},$$
$$C_2 \subset \overline{K}_2(x^*, \alpha) = \{x \mid \langle x, x^*\rangle \leqslant \alpha\}.$$

于是

$$\forall x \in C_1, \langle x, x^*\rangle \geqslant \alpha; \quad \forall x \in C_2, \langle x, x^*\rangle \leqslant \alpha.$$

故

$$\inf\{\langle x, x^*\rangle \mid x \in C_1\} \geqslant \sup\{\langle x, x^*\rangle \mid x \in C_2\}.$$

因为 C_1, C_2 至少有一个不包含在 $H(x^*, \alpha)$之中,不妨设 $C_1 \not\subset H(x^*, \alpha)$,则至少存在一个点 $x_1 \in C_1, \langle x_1, x^*\rangle > \alpha$,故

$$\sup\{\langle x, x^*\rangle \mid x\in C_1\}>\inf\{\langle x, x^*\rangle \mid x\in C_2\},$$

即(1),(2)式成立. ▌

下面的定理是定理 5.3 的一个引理.

定理 5.2 设C是 R^n 中的非空相对开凸集,M是 R^n 中不与C相交的非空仿射集. 则存在超平面H,$M\subset H$,使C包含在H一侧的开半空间中.

证明 由已知条件,$M\neq R^n$.如果M是超平面,在M的一侧的开半空间必包含C,定理的结论成立.因为否则M应与C相交,而这与已知条件矛盾(如果C包含两个相对的开半空间的点x和y,则连结x和y的线段上的某一点应在两个半空间的公共边界M上).

设M不是超平面,则$\dim M\leqslant n-2$. 下面证明总可以构造一个比M高一维、不与C相交且包含M的仿射集M'. 这样,最多进行$n-1$次相似的步骤,就可以得到具有所需性质的超平面H,从而证明了定理.

不失一般性,设$\theta\in M$,即M是一个子空间. 显然,$\theta\notin C-M$, $C\subset C-M$. 易证$C-M$也是凸集. 因为$\dim M\leqslant n-2$,故$\dim M^{\perp}\geqslant 2$, $M^{\perp}$应包含一个2维子空间P. 设$C'=P\cap(C-M)$,根据推论 4.8.1 和 4.9.2,知

$$\begin{aligned}\text{ri}C' &= P\cap \text{ri}(C-M) = P\cap(\text{ri}C-\text{ri}M)\\ &= P\cap(C-M)=C'.\end{aligned}$$

所以C'是相对开集,而$\theta\notin C-M$,故$\theta\notin C'$.

下面要做的是在P中找一条过原点但不与C'相交的直线L. 于是$M'=M+L$将是比M高一维而不与C相交的子空间. 因为,反之由$(M+L)\cap C\neq\varnothing$应推出$L\cap(C-M)\neq\varnothing$,从而与$L\cap C'=\varnothing$矛盾.

为简单计,可以认为P与R^2相同. 如果C'是-1维或0维的,直线L无疑将存在. 如果$\text{aff}\, C'$是包含θ的直线,可以取L是过θ的该直线的垂线. 如果C'是2维(因而是开集)的情形,在§8将

会证明,集合 $K=\bigcup\{\lambda C'|\lambda>0\}$是包含 C' 的最小凸锥,它是开集的并,因而也是开的且不包含 θ. 所以K是 R^2 中张角不大于π的开扇形. 这时只要将扇形的两条边界射线之一延长,将所得的直线取作L即可. ▌

定理 5.3 设 C_1,C_2是 R^n 中的非空凸集,则存在分离 C_1,C_2 的超平面的充分必要条件是 $\mathrm{ri}\,C_1\cap \mathrm{ri}C_2=\varnothing$.

证明 设 $C=C_1-C_2$. 因为$-C_2$是凸集,故C是凸集. 由推论 4.9.2,$\mathrm{ri}C=\mathrm{ri}C_1-\mathrm{ri}\,C_2$. 所以 $\mathrm{ri}\,C_1\cap\mathrm{ri}\,C_2=\varnothing$的等价条件是$\theta\notin \mathrm{ri}\,C$.

设 $\theta\notin\mathrm{ri}C$. 根据定理 5.2,存在超平面H,使 $M=\{\theta\}\subset H$,而$\mathrm{ri}C$包含在以H为边界的一个开半空间之中. 又因为$C\subset\mathrm{cl}(\mathrm{ri}C)$,故以$H$为边界的相应闭半空间也包含$C$. 于是存在一个 $x^*\in R^n$,使

$$H(x^*,0)=\{x|\langle x,x^*\rangle=0\},$$

$$0\leqslant\inf_{x\in C}\langle x,x^*\rangle=\inf_{x_1\in C_1,x_2\in C_2}\langle x_1-x_2,x^*\rangle$$

$$=\inf_{x_1\in C_1}\langle x_1,x^*\rangle-\sup_{x_2\in C_2}\langle x_2,x^*\rangle, \tag{4}$$

$$0<\sup_{x\in C}\langle x,x^*\rangle=\sup_{x_1\in C_1}\langle x_1,x^*\rangle-\inf_{x_2\in C_2}\langle x_2,x^*\rangle. \tag{5}$$

根据定理 5.1,$H(x^*,0)$分离C_1,C_2.

设 C_1,C_2可以分离. 根据定理 5.1,存在 $x^*\in R^n$,使(4),(5)式成立,则

$$C\subset D=\{x|\langle x,x^*\rangle\geqslant0\},\ \mathrm{ri}\,D\cap C\neq\varnothing.$$

由推论 4.8.2,$\mathrm{ri}C\subset\mathrm{ri}\,D$. 但$\theta\notin\mathrm{ri}\,D$,故$\theta\notin\mathrm{ri}\,C$. ▌

例 2 设

$$C_1=\{(\xi_1,\xi_2)|\xi_1>0,\xi_2\geqslant\xi_1^{-1}\},$$

$$C_2=\{(\xi_1,0)|\xi_1\geqslant0\}.$$

显然 $\mathrm{ri}\,C_1\cap\mathrm{ri}\,C_2=\varnothing$. C_1,C_2被唯一的分离超平面 $H=\{(\xi_1,0)|$

$\xi_1\in R\}$分离，$C_2\subset H$. 这个例子再一次说明分离允许两个集合之一(而不是两个)包含在分离超平面之中．见图 16.

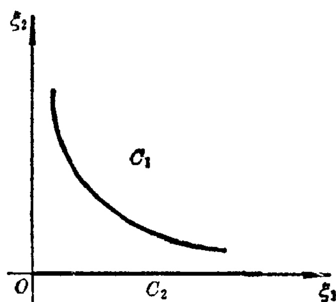


图 16

下面的结果是凸集可以分离的充分条件.

定理 5.4 设 C_1, C_2 是 R^n 中的凸集，如果 int $C_1\neq\varnothing$, $C_2\cap$ int $C_1=\varnothing$，则存在分离 C_1, C_2的超平面.

证明 首先设 C_1 是开集，即 $C_1=\text{int}\, C_1$. 则 C_1-C_2 是开凸集．因为 $C_1\cap C_2=\varnothing$, $\theta\notin C_1-C_2$, 故存在超平面 $H(a,0)=\{x|\langle a,x\rangle=0\}$满足 $\theta\in H(a,0)$, $H(a,0)\cap(C_1-C_2)=\varnothing$，且

$$C_1-C_2\subset\{x|\langle a,x\rangle>0\}. \tag{6}$$

由(6)式知 $\forall x_1\in C_1$, $\forall x_2\in C_2$时

$$\langle a,x_1-x_2\rangle>0,$$

故

$$\langle a,x_1\rangle>\langle a,x_2\rangle,\forall x_1\in C_1,\forall x_2\in C_2, \tag{7}$$

$$\inf_{x_1\in C_1}\langle a,x_1\rangle\geqslant\sup_{x_2\in C_2}\langle a,x_2\rangle, \tag{8}$$

$$\sup_{x_1\in C_1}\langle a,x_1\rangle>\inf_{x_2\in C_2}\langle a,x_2\rangle. \tag{9}$$

故超平面$H(a,\alpha)=\{x|\langle a,x\rangle=\alpha\}$分离$C_1$, C_2，其中α满足$\inf\limits_{x_1\in C_1}\langle a,x_1\rangle\geqslant\alpha\geqslant\sup\limits_{x_2\in C_2}\langle a,x_2\rangle$.

再设 C_1 不是开集，对intC_1, C_2重复上面的讨论可知，存在$a\in R^n$，有

$$\inf_{x_1\in \text{int}C_1}\langle a,x_1\rangle\geqslant\sup_{x_2\in C_2}\langle a,x_2\rangle, \tag{10}$$

$$\sup_{x_1\in \text{int}C_1}\langle a,x_1\rangle>\inf_{x_2\in C_2}\langle a,x_2\rangle. \tag{11}$$

但是 $\inf\limits_{x_1\in \mathrm{int}C_1}\langle a,x_1\rangle=\inf\limits_{x_1\in C_1}\langle a,x_1\rangle$, $\sup\limits_{x_1\in \mathrm{int}C_1}\langle a,x_1\rangle=\sup\limits_{x_1\in C_1}\langle a,x_1\rangle$,
故由(10)，(11)式知(8)，(9)式仍成立，所以超平面 $H(a,\alpha)$分离 C_1,C_2。参见图17。▌

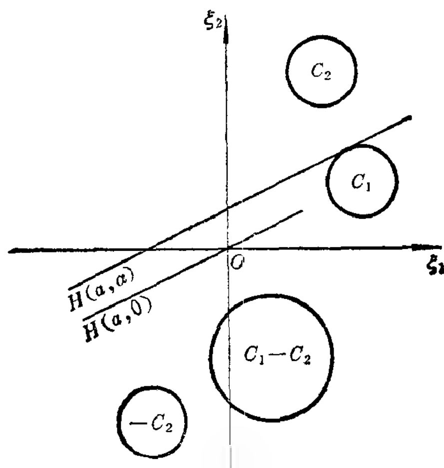


图 17

例3 在R^3中，设

$$C_1=\{(\xi_1,\xi_2,\xi_3)\mid|\xi_1|\leqslant1,|\xi_2|\leqslant1,\xi_3=0\},$$
$$C_2=\{(\xi_1,\xi_2,\xi_3)\mid\xi_1=\xi_2=0,|\xi_3|\leqslant1\}.$$

显然$\mathrm{int}C_1=\varnothing$，$C_2\cap\mathrm{int}C_1=\varnothing$，但不存在任何超平面分离 C_1,C_2. 事实上，$\mathrm{ri}C_1\cap\mathrm{ri}C_2\neq\varnothing$，故由定理 5.3，$C_1,C_2$ 不能分离是很自然的.

例 3 说明，定理 5.4 中条件$\mathrm{int}C_1\neq\varnothing$不能削弱.

因为两个凸集不相交一定可以严格分离，所以下面的结果是正确的.

定理5.5 设C_1,C_2是R^n中的凸集，$C_1\cap C_2=\varnothing$. 则存在非零向量$x^*\in R^n$，使对于$\forall x_1\in C_1,\forall x_2\in C_2$，有

$$\langle x_1,x^*\rangle\leqslant\langle x_2,x^*\rangle. \tag{12}$$

证明留给读者.

2. 超平面H强分离C_1,C_2的条件

定理5.6 设C_1,C_2是R^n中的非空凸集,则存在超平面强分离C_1,C_2的充分必要条件是存在一个$x^*\in R^n$,满足

$$\inf\{\langle x,x^*\rangle\mid x\in C_1\}>\sup\{\langle x,x^*\rangle\mid x\in C_2\}. \tag{13}$$

证明 设存在$x^*\in R^n$满足(13)式,显然,$x^*\neq\theta$. 取$\alpha\in R$,$\delta>0$$\left(\text{例如设}M=\inf\{\langle x,x^*\rangle\mid x\in C_1\},m=\sup\{\langle x,x^*\rangle\mid x\in C_2\},\alpha=\frac{M+m}{2},\delta=\frac{M-m}{2}\right)$使

$$\forall x_1\in C_1,\ \langle x_1,x^*\rangle\geqslant\alpha+\delta,\forall x_2\in C_2,\langle x_2,x^*\rangle\leqslant\alpha-\delta.$$

因为欧氏单位球B有界,故当$\varepsilon>0$充分小时,对$\forall y\in\varepsilon B$,$|\langle y,x^*\rangle|<\delta$. 则对$\forall x_1\in C_1,\forall x_2\in C_2$,有

$$\langle x_1+y,x^*\rangle=\langle x_1,x^*\rangle+\langle y,x^*\rangle\geqslant\alpha+\delta+\langle y,x^*\rangle>\alpha, \tag{14}$$

$$\langle x_2+y,x^*\rangle=\langle x_2,x^*\rangle+\langle y,x^*\rangle\leqslant\alpha-\delta+\langle y,x^*\rangle<\alpha \tag{15}$$

故由(14),(15)式知

$$C_1+\varepsilon B\subset\{x\mid\langle x,x^*\rangle>\alpha\}, \tag{16}$$

$$C_2+\varepsilon B\subset\{x\mid\langle x,x^*\rangle<\alpha\}. \tag{17}$$

所以超平面$H(x^*,\alpha)$强分离C_1,C_2.

反之,设超平面$H(x^*,\alpha)=\{x\mid\langle x,x^*\rangle=\alpha\}$强分离$C_1,C_2$,对于某个$\varepsilon>0$,(16),(17)式成立.则

$$\alpha\leqslant\inf\{\langle x,x^*\rangle+\varepsilon\langle y,x^*\rangle\mid x\in C_1,y\in B\}<\inf\{\langle x,x^*\rangle\mid x\in C_1\},$$

$$\alpha\geqslant\sup\{\langle x,x^*\rangle+\varepsilon\langle y,x^*\rangle\mid x\in C_2,y\in B\}>\sup\{\langle x,x^*\rangle\mid x\in C_2\}.$$

故(13)式成立. ▌

定理5.7 设C_1,C_2是R^n中的非空凸集. 则存在强分离C_1,C_2的超平面的充分必要条件是

$$\theta\notin\mathrm{cl}(C_1-C_2),$$

即$\inf\{|x_1-x_2| \mid x_1\in C_1, x_2\in C_2\}>0$.

证明 由定义 5.1，C_1,C_2可以强分离的等价条件是存在一个$\varepsilon>0$，使$(C_1+\varepsilon B)\cap(C_2+\varepsilon B)=\varnothing$，其中$B$是欧氏单位球. 于是$C_1,C_2$可以强分离的等价条件是存在$\varepsilon>0$，满足

$$\theta\notin(C_1+\varepsilon B)-(C_2+\varepsilon B)=C_1-C_2+2\varepsilon B. \tag{18}$$

(18)式表示存在$\varepsilon>0$，使

$$2\varepsilon B\cap(C_1-C_2)=\varnothing,$$

即$\theta\notin \mathrm{cl}(C_1-C_2)$. ▌

从例 2 可知，即使两个凸集是不相交的闭集，也不能保证可以强分离. 而问题正好出在两个集合的无界性上. 如果两个集合都是不相交的紧致凸集，则总是可以强分离的. 事实上，下面定理的条件还要宽些.

定理5.8 设C_1,C_2是R^n中的非空凸集，C_1是紧致集，C_2是闭集. 则存在强分离C_1,C_2的超平面的充分必要条件是C_1,C_2不相交.

证明 设C_1,C_2不相交. 令

$$d(C_1,C_2)=\inf\{d(x,y) \mid x\in C_1, y\in C_2\}. \tag{19}$$

因为C_1是紧致集，C_2是闭集，$C_1\cap C_2=\varnothing$，故$\delta=d(C_1,C_2)>0$. 设$S=B(\theta,\dfrac{\delta}{2})$是中心在原点，半径为$\dfrac{\delta}{2}$的开球，于是$C_1+S, C_2+S$是不相交的开凸集. 由定理 5.3，存在超平面$H$分离$C_1+S, C_2+S$. 所以超平面$H$强分离$C_1,C_2$.

反之，如果超平面H强分离C_1,C_2，则由定义 5.1，存在一个$\varepsilon>0$，使

$$(C_1+\varepsilon B)\cap(C_2+\varepsilon B)=\varnothing,$$

其中B是欧氏单位球. 显然，$C_1\cap C_2=\varnothing$. ▌

定理 5.8 中C_1,C_2是凸集的条件还可以放宽.

定理5.9 设C_1,C_2是R^n中的非空紧致集. 则存在强分离C_1,

C_2的超平面的充分必要条件是$coC_1 \cap coC_2 = \varnothing$.

证明 设$coC_1 \cap coC_2 = \varnothing$. 因为$coC_1$, coC_2仍是紧致集，由定理 5.8，存在超平面H强分离coC_1, coC_2，即存在$\varepsilon > 0$，有$(coC_1 + \varepsilon B) \cap (coC_2 + \varepsilon B) = \varnothing$，其中$B$是欧氏单位球. 因为$C_1 \subset coC_1$, $C_2 \subset coC_2$，所以有

$$(C_1 + \varepsilon B) \cap (C_2 + \varepsilon B) = \varnothing,$$

即超平面H强分离C_1, C_2.

反之，设C_1, C_2可以强分离. 由定理 5.6，存在$x^* \in R^n$, $\alpha \in R$，使

$$\inf\{\langle x_1, x^* \rangle \mid x_1 \in C_1\} > \alpha > \sup\{\langle x_2, x^* \rangle \mid x_2 \in C_2\}. \tag{20}$$

设$x = \lambda_1 x_1 + \cdots + \lambda_m x_m$，其中$\lambda_i \geqslant 0$, $i = 1, \cdots, m$, $\sum_{i=1}^{m} \lambda_i = 1$. $x_i \in C_1$, $i = 1, \cdots, m$. 则

$$\begin{aligned} \langle x, x^* \rangle &= \lambda_1 \langle x_1, x^* \rangle + \cdots + \lambda_m \langle x_m, x^* \rangle \\ &> \lambda_1 \alpha + \cdots + \lambda_m \alpha = \alpha. \end{aligned} \tag{21}$$

由(21)式知 $\forall x \in coC_1$, $\langle x, x^* \rangle > \alpha$. 类似推证可以得出，$\forall x \in coC_2$, $\langle x, x^* \rangle < \alpha$. 则

$$\inf\{\langle x_1, x^* \rangle \mid x_1 \in coC_1\} > \sup\{\langle x_2, x^* \rangle \mid x_2 \in coC_2\}.$$

所以coC_1, coC_2可以强分离，从而$coC_1 \cap coC_2 = \varnothing$. ▌

推论5.9.1 设$x \in R^n$，C是R^n中的紧致集. 则存在超平面强分离x和C的充分必要条件是$x \notin coC$.

证明 令$\{x\} = C_1$, $C = C_2$，利用定理 5.9 即可得证. ▌

推论5.9.2 设$x \in R^n$，C是R^n中的紧致集. 则存在超平面强分离x和C的充分必要条件是对于由C中最多$n+1$个点构成的每一个集合T，x与T强分离.

证明留给读者.

推论5.9.3 设C_1, C_2是R^n中的非空紧致集. 则存在超平面

强分离C_1,C_2的充分必要条件是对于由C_2中最多$n+1$个点构成的每一个集合T，C_1,T可以强分离.

证明留给读者.

§ 6 闭凸集的表示定理

这一节研究闭凸集的表示定理，它们又分为外表示定理和内表示定理. 为此，将讨论支撑超平面和面结构的有关结果.

1. 闭凸集的外表示

定义6.1 设C是R^n中的凸集. 如果闭半空间$\overline{K}(x^*,\alpha)=\{x|\langle x,x^*\rangle\leqslant\alpha\}$包含$C$，且$H(x^*,\alpha)\cap C\neq\varnothing$，其中$H(x^*,\alpha)=\{x|\langle x,x^*\rangle=\alpha\}$，则称$\overline{K}(x^*,\alpha)$是$C$的支撑半空间，而$H(x^*,\alpha)$称为$C$的支撑超平面.

由定义6.1可知，凸集C的支撑超平面$H(x^*,\alpha)$应该满足：$\forall x\in C$，$\langle x,x^*\rangle\leqslant\alpha$，且至少存在一个$x_0\in C$，$\langle x_0,x^*\rangle=\alpha$. 见图18.

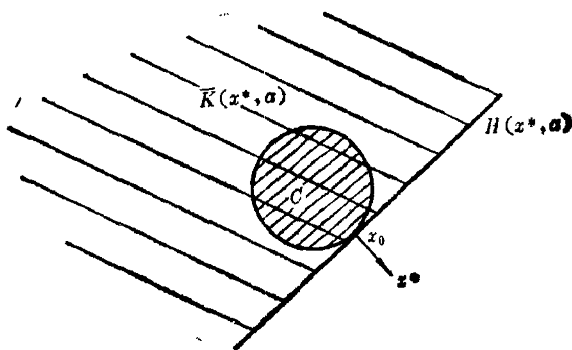


图 18

如果C是闭凸集，则定义6.1中支撑超平面的定义可以表示成：超平面$H(x^*,\alpha)$是非空闭凸集C的支撑超平面的等价条件是

$$\alpha=\max_{x\in C}\langle x,x^*\rangle, \tag{1}$$

或

$$\alpha=\min_{x\in C}\langle x,x^*\rangle. \tag{2}$$

如果对于$H(x^*,\alpha)$,(2)式成立，则对于$H(-x^*,-\alpha)$，(1)式成立。因为$H(x^*,\alpha)=H(-x^*,-\alpha)$. 故$C$的任何支撑超平面均可表示为$H(x^*,\alpha)$的形式，其中$x^*$，$\alpha$满足(1)式，而这意味着$C\subset\overline{K}(x^*,\alpha)$.

在定义6.1中，如果$C\subset H(x^*,\alpha)$，即以$H(x^*,\alpha)$为边界的两个闭半空间$\overline{K}(x^*,\alpha)=\{x|\langle x^*,\alpha\rangle\leqslant\alpha\}$，$\underline{K}(x^*,\alpha)=\{x|\langle x,x^*\rangle\geqslant\alpha\}$都是$C$的支撑半空间，则$H(x^*,\alpha)$称为非正常支撑超平面. 如果$C\not\subset H(x^*,\alpha)$，则$H(x^*,\alpha)$称为正常支撑超平面.

容易证明，当且仅当

$$\inf_{x\in C}\langle x,x^*\rangle<\max_{x\in C}\langle x,x^*\rangle, \tag{3}$$

满足$C\subset\overline{K}(x^*,\alpha)$的支撑超平面$H(x^*,\alpha)$才是正常支撑超平面.

定理6.1 设C, H分别是R^n中的非空凸集和超平面，则下面两个条件等价：

1) $H\cap\mathrm{ri}C=\varnothing$,

2) C包含在以H为边界的两个闭半空间的一个之中，但C不包含在H之中.

证明 设1)成立，因为$\mathrm{ri}C\neq\varnothing$，故可取$x_0\in\mathrm{ri}C$. 则$x_0\notin H$，所以$C$不包含在$H$之中.

假设存在点$x_1\in C$，而x_0, x_1位于超平面的两侧，则存在$x^*\in R^n$和$\alpha\in R$，使$H=H(x^*,\alpha)$，且

$$\langle x_0,x^*\rangle<\alpha<\langle x_1,x^*\rangle. \tag{4}$$

令

$$\lambda=\frac{\alpha-\langle x_0,x^*\rangle}{\langle x_1,x^*\rangle-\langle x_0,x^*\rangle},$$

则由(4)知，$0<\lambda<1$，点$x_\lambda=(1-\lambda)x_0+\lambda x_1$是线段 $\overline{x_0x_1}$ 的相对内点. 直接计算可知$\langle x_\lambda, x^*\rangle=\alpha$，故$x_\lambda\in H$. 另一方面，由定理4.5，$x_\lambda\in \mathrm{ri}C$，所以$x_\lambda\in H\cap \mathrm{ri}C$. 这与1)矛盾. 所以$C$包含在以$H$为边界的一个闭半空间之中. 事实上，这个闭半空间就是包含x_0的闭半空间. 故2)成立.

设2)成立. 假设存在点$x\in H\cap \mathrm{ri}C$. 因为C不包含在H之中，故存在点$x_1\in C\backslash H$，从而由定理4.6，存在点$x_2\in C$及$0<\lambda<1$，使

$$x=(1-\lambda)x_1+\lambda x_2.$$

另一方面，由2)，可以找到$x^*\in R^n$，$\alpha\in R$，使$H=H(x^*,\alpha)$，$\overline{K}(x^*,\alpha)=\{x\mid\langle x, x^*\rangle\leqslant\alpha\}$，且$C\subset\overline{K}(x^*,\alpha)$. 显然$\langle x_1,x^*\rangle<\alpha$，$\langle x_2,x^*\rangle\leqslant\alpha$，所以

$$\begin{aligned}\langle x,x^*\rangle&=\langle(1-\lambda)x_1+\lambda x_2,x^*\rangle\\&=(1-\lambda)\langle x_1,x^*\rangle+\lambda\langle x_2,x^*\rangle\\&<(1-\lambda)\alpha+\lambda\alpha=\alpha.\end{aligned}\tag{5}$$

但因为$x\in H$，而有$\langle x,x^*\rangle=\alpha$。这与(5)式相矛盾. 所以$H\cap \mathrm{ri}C=\varnothing$，1)成立. ▌

推论6.1.1 设C是R^n中的闭凸集，H是C的支撑超平面. 则H是正常超平面的充分必要条件是$H\cap \mathrm{ri}C=\varnothing$.

证明留给读者.

下面讨论支撑超平面存在的条件.

定理6.2 设C，D是R^n中的凸集，$D\subset C$. 则C存在包含D的正常支撑超平面的充分必要条件是$D\cap \mathrm{ri}C=\varnothing$.

证明 因为$D\subset C$，C的包含D的正常支撑超平面和分离D与C的超平面相同. 由定理5.3，D，C存在分离超平面的等价条件是$\mathrm{ri}D\cap \mathrm{ri}C=\varnothing$.

如果$D\cap \mathrm{ri}C=\varnothing$，则$\mathrm{ri}D\cap \mathrm{ri}C=\varnothing$.

如果$\mathrm{ri}D\cap \mathrm{ri}C=\varnothing$，但因$D\subset C$，根据推论4.8.2，知$D\cap \mathrm{ri}C=$

$\varnothing$。综上述，$\mathrm{ri}D\cap\mathrm{ri}C=\varnothing$的等价条件是$D\cap\mathrm{ri}C=\varnothing$. ▌

定理6.3 设C是R^n中的闭凸集，x是$\mathrm{rb}C$中的任意点. 则C存在过x的正常支撑超平面.

证明 当$\dim C=-1$或0时，结论明显成立，故设$\dim C\geqslant1$.

因为$\mathrm{ri}C$是相对开凸集，$\{x\}$是仿射集，由定理5.2，在$\mathrm{aff}C$中存在超平面H'，使

$$x\in H',\quad H'\cap\mathrm{ri}C=\varnothing.$$

显然，可以找到R^n中的超平面H，使$H\cap\mathrm{aff}C=H'$(如果$\mathrm{aff}C=R^n$，则$H=H'$). 当然也有

$$x\in H,\quad H\cap\mathrm{ri}C=\varnothing.$$

所以由定理6.1，H是C的正常支撑超平面. ▌

定理 6.4 设C是R^n中的闭集，$\mathrm{int}C\neq\varnothing$. 如果$C$存在通过每一个边界点的支撑超平面，则$C$是凸集.

证明 仅在$n\geqslant2$时证明定理，$n=1$时的证明留作练习.

如果$C=R^n$，C当然是凸集，故可设$C\neq R^n$. 由此设$x\in R^n$，$x\notin C$. 如果$y\in\mathrm{int}C$，则在C的边界上存在$z\in\mathrm{ri}\overline{xy}$. 所以$C$在$z$的支撑超平面$H$不包含$x$，因为否则这个超平面将包含$C$的内点$y$，而这是不可能的. 故以$H$为边界且包含$y$的闭半空间不包含$x$但包含$C$. 因为$x$是不在$C$中的任意点，故包含$C$的闭半空间的交不包含不是$C$中的点. 另一方面，包含$C$的闭半空间的交一定包含$C$. 所以$C$是包含$C$的全体闭半空间的交. 因为闭半空间是凸集，由定理3.3，C是凸集. ▌

结合定理6.3和定理6.4，就可以得出具有非空内部的闭凸集的一个性质.

定理 6.5 设C是R^n中的闭集，$\mathrm{int}C\neq\varnothing$. 则$C$是凸集的充分必要条件是在$C$的每一个边界点均存在$C$的支撑超平面.

下面的定理称为闭凸集的外表示定理.

定理 6.6 R^n中的闭凸集C是包含它的全体闭半空间的交.

证明 当 $\dim C=-1$ 或0时，或 $C=R^n$ 时，定理的结论是明显的.

设 $\dim C\geqslant 1$. $C\neq R^n$. 与定理6.4的证明完全相似，可以证明，包含 C 的全体闭半空间的交不包含不是 C 中的点. 另一方面，C 包含在包含 C 的全体闭半空间的交之中，所以 C 是包含它的全体闭半空间的交. ▎

推论 6.6.1 设 $S\subset R^n$，则 $\mathrm{cl}(\mathrm{co}S)$ 是包含 S 的全体闭半空间的交.

证明留给读者.

2. 闭凸集的内表示

为了解决闭凸集的内表示问题，我们首先研究闭凸集的面结构问题. 最后将证明，如果 C 是紧致凸集，则它可以表示为其零维面的凸包，这就是闭凸集的内表示形式.

定义 6.2 设 C 是 R^n 中的闭凸集，$F\subset C$，F 是凸集. 如果对于 $\forall x\in C$，$\forall y\in C$，$x\neq y$，在 $\mathrm{ri}\overline{xy}\cap F$ 非空时一定有 $\overline{xy}\subset F$，则称 F 是 C 的面.

空集 $\varnothing$ 和闭凸集 C 本身也是 C 的面，称它们为非正常面，而其他所有的面均称为正常面.

定义 6.3 设 C 是 R^n 中的闭凸集，$x\in C$. 如果 $\{x\}$ 是 C 的面，则称 x 是 C 的极点. C 的全体极点的集合用 $\mathrm{ext}\,C$ 表示.

由定义6.3可知，如果 x 是 C 的一个极点，则 x 不是 C 中任何一条线段的相对内点，即 x 不能表示成 C 中任何元素的凸组合. 显然，极点的这个定义与定义3.7是等价的.

定义 6.4 设 F 是 C 的一个面，$\dim F=k$，则称 F 是 k-面. 称 C 的极点是0-面. 如果 $0\leqslant\dim F=\dim C-1$，则称 F 是超面.

定理 6.7 设 C 是 R^n 中的闭凸集，$\dim C\geqslant 1$，H 是 C 的一个正常支撑超平面，则 $F=H\cap C$ 是 C 的正常面 .

证明 首先，因为 H 是 C 的正常支撑超平面，故 F 是 C 的正常

非空的凸子集.

设x,y是使$\mathrm{ri}\overline{xy}\cap F\neq\varnothing$的$C$中的两点，则存在一个$\lambda$，$0<\lambda<1$，使$(1-\lambda)x+\lambda y\in H$.

由已知条件，可以设$H=H(x^*,\alpha)$，$C\subset\overline{K}(x^*,\alpha)$，其中$x^*\in R^n$，$\alpha\in R$. 故

$$\langle x,x^*\rangle\leqslant\alpha,\langle y,x^*\rangle\leqslant\alpha,\langle(1-\lambda)x+\lambda y,x^*\rangle=\alpha. \tag{6}$$

所以有$\langle x,x^*\rangle=\langle y,x^*\rangle=\alpha$. 即$x\in H,y\in H$. 但$C$，$H$是凸集，所以$\overline{xy}\subset F$，$F$是$C$的正常面. ▌

我们称定理6.7中的面$F=H\cap C$是C的正常暴露面. 为了方便，对于任意闭凸集C（包括-1维和0维的在内），称空集$\varnothing$和C是C的非正常暴露面. 对于$x\in C$，如果$\{x\}$是C的暴露面，则称x是C的暴露点. C的全体暴露点的集合用$\exp C$表示. 显然，$\exp C\subset\mathrm{ext}C$.

例 1 图19中C是R^2中两个圆C_1，C_2的凸包. 则C的面是线段$\overline{x_1x_3}$，$\overline{x_2x_4}$及分别过x_1,x_5,x_2和x_3,x_6，x_4的开半圆弧；线段$\overline{x_1x_3}$，$\overline{x_2y_4}$是C的1-面，且是暴露面；x_1，x_2，x_3，x_4及上述两个开半圆弧上的点都是极点(0-面)；两个开半圆弧上的点都是暴露点，而x_1、x_2，x_3，x_4都不是暴露点，因为过这些点的C的支撑超平面也应包含线段$\overline{x_1x_3}$或$\overline{x_2x_4}$.

例1也说明非暴露面是存在的.

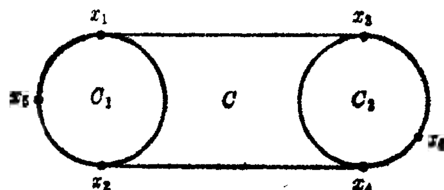


图 19

下面讨论面的一些性质.

定理 6.8 R^n中的闭凸集C的每一个面F都是闭集.

证明　当 $\dim F=-1$ 或 0 时，结论是明显的．故设 $\dim F\geqslant 1$．

取 $\forall x\in \mathrm{cl}F, x_0\in \mathrm{ri}F$．不妨设 $x\neq x_0$，因为反之，在 $x=x_0$ 时，$x\in F$，正是我们要证明的事实．

由定理 4.5，$(1-\lambda)x_0+\lambda x\in \mathrm{ri}F$，$0\leqslant\lambda<1$．特别地，$0<\lambda<1$ 时，$\{(1-\lambda)x_0+\lambda x\,|\,0<\lambda<1\}\cap F\neq\varnothing$．因为 F 是 C 的面，由面的定义 6.2 知，线段 $\overline{x_0x}$ 也包含在 F 中，$x\in F$．故 F 是闭集．▌

定理 6.9　设 F 是 R^n 中闭凸集 C 的面，$G\subset F$，当且仅当 G 是 F 的面时，G 是 C 的面．

证明　如果 G 是 C 的面，则由定义直接得到，G 也是 F 的面．

相反，设 G 是 F 的面，x,y 是 C 中满足

$$\{(1-\lambda)x+\lambda y\,|\,0<\lambda<1\}\cap G\neq\varnothing \tag{7}$$

的点．因为 $G\subset F$，则 $\{(1-\lambda)x+\lambda y\,|\,0<\lambda<1\}\cap F\neq\varnothing$．而因为 F 是 C 的面，故 $x\in F, y\in F$．同时 G 是 F 的面，于是又得到，$x\in G, y\in G$，即 $\overline{xy}\subset G$．所以 G 是 C 的面．▌

例 2　在例 1 中，x_1 是面 $\overline{x_1x_3}$ 的暴露点，而 $\overline{x_1x_3}$ 是 C 的暴露面，但 x_1 不是 C 的暴露点．

例 2 说明在定理 6.9 中，如果将"面"换成"暴露面"，则定理的充分性不再成立．

定理 6.10　设 F 是 R^n 中闭凸集 C 的面，$F\neq C$．则 $F\subset \mathrm{rb}C$．

证明　当 $\dim C=-1$ 或 0 时，定理结论显然成立．故设 $\dim C\geqslant 1$．

假设 F 是 C 的面，且存在 $x\in F$，使 $x\in \mathrm{ri}C$．

设 y 是 C 中的任意点，如果 $y=x$，则 $y\in F$，这正是我们要证明的事实．故不妨设 $y\neq x$．由定理 4.6，在 C 中存在点 z，使 $x\in\{(1-\lambda)y-\lambda z\,|\,0<\lambda<1\}$．因 $x\in F$，而 F 是 C 的面，故 $y\in F$．由 y 在 C 中的任意性知 $F=C$．这与条件 $F\neq C$ 矛盾．故 $F\subset$

$\mathrm{rb}C$. ▌

推论 6.10.1 设 F,G 是 R^n 中闭凸集 C 的面, $G\subset F$, $G\neq F$, 则 $G\subset \mathrm{rb}F$.

证明留给读者.

推论 6.10.2 设 F,G 是 R^n 中闭凸集 C 的面, $G\subset F$, $G\neq F$, 则 $\dim G<\dim F$.

证明 由推论 6.10.1, $G\subset \mathrm{rb}F$, 故利用推论 4.7.1 的结论即得证. ▌

对于闭凸集 C 的任何子集 M, 存在 C 的包含 M 的最小面, 这个最小面就是包含 M 的所有面的交. 定理 6.10 表明, 如果 M 包含 $\mathrm{ri}C$ 中的点, 那么包含 M 的最小面就是 C 本身.

定理 6.11 设 C 是 R^n 中的闭凸集, $x\in C$, F 是 C 的面, $x\in F$. 则当且仅当 $x\in \mathrm{ri}F$ 时, F 是包含 x 的最小面.

证明 设 $x\in \mathrm{ri}F$. 如果还存在 C 的面 F', $x\in F'$, $F'\subset F$, 则 $F'\cap \mathrm{ri}F\neq\varnothing$, 由推论 6.10.1, $F'=F$. 所以 F 是包含 x 的最小面.

假设 $x\in \mathrm{rb}F$. 由定理 6.3, 存在 F 的面 G, 且 $x\in G$, $G\neq F$ (由暴露面定义知, G 也可认为是暴露面). 而由定理 6.9, G 也是 C 的面, 所以 F 不是包含 x 的最小面. ▌

推论 6.11.1 设 C 是 R^n 中的闭凸集. 则对 $\forall x\in C$, C 存在唯一的面 F, 使 $x\in \mathrm{ri}F$.

证明 包含 x 的 C 的最小面即是满足推论要求的面. ▌

下面两个定理讨论暴露面的性质.

定理 6.12 R^n 中的闭凸集 C 的任何超面 F 都是暴露面.

证明 由超面的定义知 $\dim F\geqslant 0$, 故由定理 4.4, 存在 $x\in \mathrm{ri}F$. 所以由定理 6.11 知, F 是 C 的包含 x 的最小面.

另一方面, 由定理 6.3 知, 存在 C 的暴露面 G, 满足 $x\in G$. 因而 $F\subset G\subset C$, 且 $G\neq C$. 故由推论 6.10.2 知

$$\dim C-1=\dim F\leqslant\dim G<\dim C \tag{8}$$

由此可知, $\dim G=\dim F$. 再利用推论 6.10.2, 得 $F=G$, 从而 F 是暴露面 .▌

定理 6.13 设C是 R^n 中的闭凸集, $\{F_i|i\in I\}$是C的任意一个暴露面族. 则 $F=\bigcap_{i\in I}F_i$ 也是C的暴露面, 其中 I 是任意指标集.

证明 如果F是空集$\varnothing$或C本身, 定理的结论明显成立. 故设F是正常暴露面 F_i 的非空交集, $i\in I$.

首先研究I是有限指标集的情形, 不妨设 $I=\{1,\cdots,m\}$. 对于每一个 $F_i, i\in I$, 由暴露面的定义, 存在支撑超平面 $H(x_i^*,\alpha_i)=\{x|\langle x,x_i^*\rangle=\alpha_i\}$及闭半空间 $\overline{K}(x_i^*,\alpha_i)=\{x|\langle x,x_i^*\rangle\leqslant\alpha_i\}$, $x_i^*\in R^n$, $\alpha_i\in R$, 使

$$F_i=H(x_i^*,\alpha_i)\cap C, \tag{9}$$

$$C\subset\overline{K}(x_i^*,\alpha_i). \tag{10}$$

不失一般性, 可以设 $\theta\in\text{int}C$. 所以 θ 是所有 $\overline{K}(x_i^*,\alpha_i)$的内点; 且 $\alpha_i>0, i\in I$. 令

$$y_i^*=\alpha_i^{-1}x^*,\quad i=1,\cdots,m.$$

于是(9), (10)可以改写成

$$F_i=H(y_i^*,1)\cap C, \tag{11}$$

$$C\subset\overline{K}(y_i^*,1). \tag{12}$$

令

$$y_0^*=\frac{1}{m}\sum_{i=1}^{m}y_i^*.$$

对于 $\forall x\in C$, 有

$$\langle x,y_0^*\rangle=\langle x,\frac{1}{m}\sum_{i=1}^{m}y_i^*\rangle$$

$$=\sum_{i=1}^{m}\frac{1}{m}\langle x,y_i^*\rangle\leqslant\sum_{i=1}^{m}\frac{1}{m}\cdot 1=1. \tag{13}$$

由(13),得 $C\subset\overline{K}(y_0^*,1)$. (13)中的等号成立的等价条件是 $x\in H(y_i^*,1),i=1,\cdots,m$. 这表示

$$F=H(y_0^*,1)\cap C. \tag{14}$$

所以 F 是 C 的暴露面.

再研究 I 是无限指标集的情形. 根据上面所证的结果, 只要证明存在 $i_1,\cdots,i_m\in I$, 使 $\bigcap\limits_{j=1}^{m}F_{i_j}=F$ 就可以了.

在 I 中任取一个元素作为 i_1,如果 $F=F_{i_1}$, 结论得证. 如果 $F\subset F_{i_1}$,且 $F\neq F_{i_1}$,则对某一个 $i_2\in I$,有

$$F\subset F_{i_1}\cap F_{i_2}\subset F_{i_1}, \tag{15}$$

$$F_{i_1}\cap F_{i_2}\neq F_{i_1}. \tag{16}$$

由推论 6.10.2, 知 $\dim(F_{i_1}\cap F_{i_2})<\dim F_{i_1}$. 如果 $F=F_{i_1}\cap F_{i_2}$,结论得证. 如果 $F\subset F_{i_1}\cap F_{i_2}$, 且 $F\neq F_{i_1}\cap F_{i_2}$,则对于某一个 $i_3\in I$,有

$$F\subset F_{i_1}\cap F_{i_2}\cap F_{i_3}\subset F_{i_1}\cap F_{i_2}, \tag{17}$$

$$F_{i_1}\cap F_{i_2}\cap F_{i_3}\neq F_{i_1}\cap F_{i_2}. \tag{18}$$

同样,由推论 6.10.2, 知 $\dim(F_{i_1}\cap F_{i_2}\cap F_{i_3})<\dim(F_{i_1}\cap F_{i_2})$. 如果 $F=F_{i_1}\cap F_{i_2}\cap F_{i_3}$, 结论得证. 否则, 与上面类似的过程可以继续下去. 因为每经过一步,维数至少降低一维,则在有限步后这个过程一定会结束且得到 $i_1,\cdots,i_m\in I$,使 $F=\bigcap\limits_{j=1}^{m}F_{i_j}$. ▌

最后, 讨论紧致凸集的内表示形式.

我们知道, 闭半空间和仿射集是没有极点的闭凸集. 对于紧致凸集, 情形就不同了, 下面的定理说明, 紧致凸集是由其极点"张"成的.

定理 6.14 设 C 是 R^n 中的紧致凸集，则

$$C=\mathrm{co}(\mathrm{ext}C). \tag{19}$$

证明 因为 $C\supset \mathrm{ext}C$，且 C 是凸集，故 $C\supset \mathrm{co}(\mathrm{ext}C)$ 显然成立。要证 $C\subset \mathrm{co}(\mathrm{ext}\,C)$ 也成立，从而(19)也是成立的。

对 C 的维数进行归纳证明。在 $\dim C=-1$，0 或 1 时，结论显然是成立的。假设上述包含关系对维数小于 k 的紧致凸集都是正确的，$k\geqslant 2$。研究 k 维紧致凸集 C。设 $\forall x\in C$，由定理 3.5，只要证明 x 可以表示成 C 的极点的凸组合就可以了。

如果 x 是 C 的极点，结论显然成立。如果 x 不是极点，则在 C 中存在一条线段，使 x 是这条线段的相对内点，由定理 4.6，这条线段可以延长而不超出 C。故可以找到 y_0，$y_1\in \mathrm{rb}C$，使 $x\in \mathrm{ri}\{(1-\lambda)y_0+\lambda y_1|0\leqslant\lambda\leqslant 1\}$。令 F_0，F_1 分别是包含 y_0，y_1 的 C 的最小面。则由推论 6.11.1，F_0，F_1 是 C 的正常面。由定理 6.8，F_0，F_1 是紧致凸集。故由推论 6.10.2，它们的维数均小于 k。于是由归纳假设，存在 F_0 的极点 x_{01}，…，x_{0p}，使 y_0 是它们的凸组合，也存在 F_1 的极点 x_{11}，…，x_{1q}，使 y_1 是它们的凸组合。因为 x 是 y_0，y_1 的凸组合，所以 x 也是 $x_{01},\cdots,x_{0p}$，$x_{11},\cdots,x_{1q}$ 的凸组合。而由定理 6.9，x_{01}，…，x_{0p}，$x_{11},\cdots,x_{1q}$ 也是 C 的极点，所以 x 可以表示成 C 的极点的凸组合。▌

定理 6.15 设 C 是 R^n 中的紧致凸集，M 是 C 的子集。则下面两个条件等价：

1) $C=\mathrm{co}M$，

2) $\mathrm{ext}C\subset M$。

证明 假设在 C 中存在一个极点 x，$x\notin M$，则 $M\subset C\setminus\{x\}$。由极点的定义，$C\setminus\{x\}$ 是凸集，故 $\mathrm{co}M\subset C\setminus\{x\}$. $C\neq \mathrm{co}M$. 所以由1)推出2)。

设 $\mathrm{ext}C\subset M$，因为由定理6.14，$C=\mathrm{co}(\mathrm{ext}C)$，故

$$C=\mathrm{co}(\mathrm{ext}C)\subset \mathrm{co}M\subset C. \tag{20}$$

所以 $C=\mathrm{co}M$，由2)推出1). ▎

推论6.15.1 设 C 是 R^n 中的紧致凸集，$\dim C=n$. 则 C 的每一点均可表示成 C 的至多 $n+1$ 个极点的凸组合.

证明留给读者.

§7 配极

对偶概念在凸性理论中具有重要的作用. 在这一节，我们讨论一类特殊的对偶概念——配极. 将会看到，对于 R^n 中的每一个子集 M，都有一个 R^n 中称为 M 的配极的集合 M° 与之对应. 如果 C 是包含原点作为内点的紧致凸集，它的配极 C° 也具有同样的性质且 C 是 C° 的配极.

1. 配极的概念与一般性质

定义7.1 设 M 是 R^n 中的集合. 称

$$\begin{aligned}M^\circ&=\{y\in R^n|\langle x,\ y\rangle\leqslant 1,\ x\in M\}\\&=\{y\in R^n|\sup_{x\in M}\langle x,y\rangle\leqslant 1\}\end{aligned} \tag{1}$$

为 M 的配极.

由定义 7.1 知，(1)式也等价于

$$M^\circ=\bigcap_{x\in M}\overline{K}(x,\ 1), \tag{2}$$

其中 $\overline{K}(x,\ 1)=\{y\in R^n|\langle y,\ x\rangle\leqslant 1\}$. 因为当且仅当 $x\in\overline{K}(y,\ 1)$ 时 $y\in\overline{K}(x,\ 1)$，故由(2)知，当且仅当 $y\in M^\circ$ 时 $M\subset\overline{K}(y,\ 1)$.

例 1 设 $M=\{x\}$，$x\neq\theta$. 则 M° 是闭半空间

$$\overline{K}(x,\ 1)=\{y|\langle x,\ y\rangle\leqslant 1\},$$

θ 是 M° 的内点.

例 2 设 $M=\{\theta\}$，则 $M^\circ=R^n$，这是因为对于 $\forall y\in R^n$，有

$$\langle\theta,\ y\rangle=0<1.$$

例 3 设$B(\theta, r)$是以原点为中心，r为半径的闭球. 下面证明

$$[B(\theta, r)]^{\circ} = B(\theta, \frac{1}{r}) .$$

由线性代数知识知，当x位于以θ为端点、经过y的射线上时，有

$$\langle x, y\rangle = |x||y|.$$

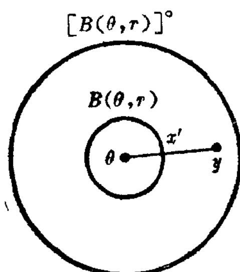


图 20

设$y \in [B(\theta, r)]^{\circ}$，$x'$是以$\theta$为端点、经过$y$的射线与$B(\theta, r)$边界的交点(见图20)，则由配极的定义，有

$1 \geqslant \langle x', y\rangle = |x'||y| = r|y|$.

故$|y| \leqslant \frac{1}{r}$，$y \in B(\theta, \frac{1}{r})$. 即$[B(\theta, r)]^{\circ} \subset B(\theta, \frac{1}{r})$.

反之，设$z \in B(\theta, \frac{1}{r})$，则对$\forall x \in B(\theta, r)$，有

$$\langle x, z\rangle = |x||z|\cos\gamma \leqslant r \cdot \frac{1}{r} \cdot 1 = 1,$$

其中γ是$\overrightarrow{\theta x}$和$\overrightarrow{\theta z}$之间的夹角，所以$z \in [B(\theta, r)]^{\circ}$，即$[B(\theta, r)]^{\circ} \supset B(\theta, \frac{1}{r})$. 最后得到$[B(\theta, r)]^{\circ} = B(\theta, \frac{1}{r})$.

定理 7.1 设M，M_1，M_2是R^n中的非空集合. 则有下列结论成立：

1) $(M_1 \cup M_2)^{\circ} = M_1^{\circ} \cap M_2^{\circ}$，

2) 如果$M_1 \subset M_2$，则$M_2^{\circ} \subset M_1^{\circ}$，

3) 如果$\lambda > 0$，则$(\lambda M)^{\circ} = \frac{1}{\lambda} M^{\circ}$，

4) M°是包含原点的闭凸集.

证明 1) 由配极的定义，有

$$(M_1\cup M_2)^\circ=\{y\mid\langle x,\ y\rangle\leqslant 1,\ x\in M_1\cup M_2\}$$
$$=\{y\mid\langle x,\ y\rangle\leqslant 1,\ x\in M_1\}\cap\{y\mid\langle x,\ y\rangle\leqslant 1,\ x\in M_2\}$$
$$=M_1^\circ\cap M_2^\circ.$$

2) 如果 $M_1\subset M_2$，则 $M_2=M_1\cup(M_2\setminus M_1)$. 于是由1) 知 $M^\circ_2=M_1^\circ\cap(M_2\setminus M_1)^\circ\subset M_1^\circ$.

3) 如果 $\forall x\in(\lambda M)^\circ$，则由配极的定义，对于 $\forall\lambda y\in\lambda M$，有$\langle\lambda y,\ x\rangle\leqslant 1$，即对于 $\forall y\in M$，有$\langle y,\ \lambda x\rangle\leqslant 1$. 而这表示 $\lambda x\in M^\circ$，$x\in(\frac{1}{\lambda})M^\circ$.

类似可证其逆也成立.

4) 由例1，例2可知，R^n 中一个点构成的集合的配极是 R^n 或闭半空间. 则由1)（1)的结论可以在任意个集合的情形成立)，有

$$M^\circ=[\bigcup_{x\in M}\{x\}]^\circ=\bigcap_{x\in M}\{x\}^\circ.$$

所以 M° 是闭半空间族的交，故它是闭凸集. 因为每一个$\{x\}^\circ$包含原点，所以 M°包含原点. ▌

定理 7.2 设$M\subset R^n$.

1) 如果 M 有界，则 θ 是 M° 的内点，

2) 如果 θ 是 M 的内点，则 M° 有界.

证明 1) 设 M 有界，则存在 $r>0$，使 $M\subset B(\theta,\ r)$. 由定理7.1的 2)知，$B(\theta,\ \frac{1}{r})\subset M^\circ$，而这表示 θ 是 M° 的内点.

2) 如果 θ 是 M 的内点，即存在 $r>0$，使$B(\theta,\ r)\subset M$. 同样由定理 7.1 的2)知，$M^\circ\subset B(\theta,\ \frac{1}{r})$. 这表示 M° 是有界的. ▌

定义 7.2 设M是 R^n 中的集合，M°是M的配极. 称$M^{\circ\circ}=$

$(M^{*})^{*}$为M的双配极.

定理 7.3 设M是 R^n 中的集合. 则

$$M^{**}=\text{clco}(\{\theta\}\cup M), \tag{3}$$

即 M^{**} 是包含 θ 和 M 的最小闭凸集.

证明 由(2)式及其后的说明，有

$$M^{**}=\bigcap_{y\in M^{*}}\overline{K}(y,\ 1)=\bigcap_{M\subset\overline{K}(y,1)}\overline{K}(y,\ 1). \tag{4}$$

因为 $M\subset\overline{K}(y,\ 1)$，$\theta\in\overline{K}(y,\ 1)$，故由(4)式知，$M^{**}$ 是包含θ和M的闭凸集. 但clco$(\{\theta\}\cup M)$是包含 θ 和 M 的最小闭凸集，所以，$M^{**}\supset\text{clco}(\{\theta\}\cup M)$.

设 $z\notin\text{clco}(\{\theta\}\cup M)$. 要证明 $z\notin M^{**}$，从而 M^{**} 不包含不是 clco$(\{\theta\}\cup M)$的点，故 $M^{**}\subset\text{clco}(\{\theta\}\cup M)$，最后得到 (4) 成立. 而要证明 $z\notin M^{**}$，根据(4)，只要找到一个包含 M 而不包含 z 的闭半空间$\overline{K}(u,\ 1)$即可.

根据定理6.6，存在 clco$(\{\theta\}\cup M)$的支撑半空间$\overline{K}(y,\ \alpha)$，且 $z\notin\overline{K}(y,\ \alpha)$，则

$$\max\{\langle x,\ y\rangle\,|\,x\in\text{clco}(\{\theta\}\cup M)\}=\alpha<\langle z,\ y\rangle. \tag{5}$$

因为 $\theta\in\text{clco}(\{\theta\}\cup M)$，故 $\alpha\geqslant 0$，所以存在 $\beta>0$，满足

$$\max\{\langle x,\ y\rangle\,|\,x\in\text{clco}(\{\theta\}\cup M)\}\leqslant\beta<\langle z,\ y\rangle. \tag{6}$$

令 $u=\dfrac{1}{\beta}y$，则由(6)式得

$$\max\{\langle x,\ u\rangle\,|\,x\in\text{clco}(\{\theta\}\cup M)\}\leqslant 1<\langle z,\ u\rangle. \tag{7}$$

由(7)式知，$M\subset\overline{K}(u,\ 1)$且 $z\notin\overline{K}(u,\ 1)$. 即$\overline{K}(u,\ 1)$就是所求的闭半空间. ▌

由定理7.2和定理7.3直接得出下面的推论.

推论 7.3.1 设 C 是 R^n 中包含原点为其内点的紧致凸集，则 C^{*} 也是包含原点为其内点的紧致凸集，且 $C^{**}=C$.

证明留给读者.

在推论7.3.1的条件下，由 $\theta\in\mathrm{int}C$ 知 C 的每一个支撑超平面都是正常的，且对于某一个唯一的 $y\in R^n\setminus\{\theta\}$，它具有 $H(y, 1)$ 的形式．于是 $C\subset\overline{K}(y,1)$，而这表示 $y\in C^\circ$．

定理 7.4 设 C 是 R^n 中包含原点为其内点的紧致凸集，对于 $\forall y\in R^n$，下面两个条件等价：

1) $H(y, 1)$是 C 的支撑超平面，

2) $y\in\mathrm{bd}C^\circ$，

其中 $\mathrm{bd}C^\circ$ 表示 C° 的边界．对偶地有，对于 $\forall x\in R^n$，下面两个条件也是等价的：

3) $H(x, 1)$是 C° 的支撑超平面，

4) $x\in\mathrm{bd}C$．

证明 设1)成立，则 $y\in C^\circ$ 且

$$\sup_{x\in C}\langle x, y\rangle=1. \tag{8}$$

假设 $y\in\mathrm{int}C^\circ$，则对于某一个 $\lambda>1$，有 $\lambda y\in C^\circ$．故由配极的定义得

$$\sup_{x\in C}\langle x, \lambda y\rangle\leqslant 1, \tag{9}$$

从而得到 $\sup\limits_{x\in C}\langle x, y\rangle\leqslant\frac{1}{\lambda}<1$．这与(8)式矛盾．所以 $y\in\mathrm{bd}C^\circ$，即2)成立．

反之，设2)成立．显然 $y\in C^\circ\setminus\{\theta\}$，故由配极的定义知

$$0<\sup_{x\in C}\langle x, y\rangle\leqslant 1. \tag{10}$$

假设(10)式右边是严格的不等式，则对于某一个 $\lambda>1$，下面的等式

$$\sup_{x\in C}\langle x, \lambda y\rangle=1$$

成立，而这表示 $\lambda y\in C^\circ$．因为 $\theta\in\mathrm{int}C^\circ$，$y$在连结 θ 与 λy 的线

段中，根据定理 4.5，$y\in \text{int}C^*$，这与条件矛盾，所以有

$$\sup_{x\in C}\langle x, y\rangle=1. \tag{11}$$

最后，由 C 的紧致性及线性函数的连续性，(11)式中的上确界可以达到，故 $H(y, 1)$是 C 的支撑超平面，即1)成立.

因为3)，4)中，C，C^* 是完全对称的角色，且 $C=C^{**}$，所以由1)，2)的等价性得出3)，4)的等价性. ▌

推论 7.4.1 设 C 是 R^n 中包含原点为其内点的紧致凸集则对于 $\forall x\in R^n$，$\forall y\in R^n$，下面四个条件等价：

1) $H(y, 1)$是 C 在 x 的支撑超平面，

2) $H(x, 1)$是 C^* 在 y 的支撑超平面，

3) $\langle x, y\rangle=1$，$x\in \text{bd}C$，$y\in \text{bd}C^*$，

其中 $\text{bd}C$，$\text{bd}C^*$ 分别表示 C，C^* 的边界.

4) $\langle x, y\rangle=1$，$x\in C$，$y=C^*$.

证明 由定理7.4的1)，2)的等价性推出1)，3)的等价性.

由定理7.4的3)，4)的等价性推出2)，3)的等价性.

显然，由3)推出4)的成立. 为了得到定理的结论，只要证明4)推出1)的成立即可.

设4)成立. 因为 $y\in C^*$，故 $C\subset\overline{K}(y,1)$. 由$\langle x,y\rangle=1$可知，$x\in H(y, 1)$. 因为 $x\in C$，所以 $H(y, 1)$是 C 在 x的支撑超平面，即1)成立. ▌

2. 共轭面

设 C 是 R^n 中包含原点为其内点的紧致凸集，F 是 C 的暴露面，G 是 C^* 的暴露面. 令

$$F^{\triangle}=\{y\in C^*|\langle x, y\rangle=1, x\in F\}, \tag{12}$$

$$G^{\triangle}=\{x\in C|\langle x, y\rangle=1, y\in G\}. \tag{13}$$

集合 $F^{\triangle}$ 是存在的，因为如果 F 是 C 的正常暴露面，则由推论7.4.1中1)和4)的等价性，当且仅当 $H(y, 1)$是 C 的包含 F 的

支撑超平面时，$y\in F^{\triangle}$. 同样，$G^{\triangle}$ 也是存在的.

定理 7.5 设 C 是 R^n 中包含原点为其内点的紧致凸集，F 是 C 的正常暴露面. 则由(12)式定义的 $F^{\triangle}$ 是 C° 的正常暴露面. 对于 C° 的正常暴露面 G，相似的结论也成立.

证明 只证明第一部分，第二部分证明相似.

由(12)式，有

$$F^{\triangle}=\bigcap_{x\in F}(C^{\circ}\cap H(x,1)). \tag{14}$$

因为面 F 是正常的，$\forall x\in F$ 是 $\mathrm{bd}C$ 中的点，故由定理 7.4 中 3) 和 4) 的等价性知 $H(x,1)$ 是 C° 的支撑超平面. 所以，每一个集合 $C^{\circ}\cap H(x,1)$ 是 C° 的正常暴露面. 故根据定理 6.13，$F^{\triangle}$ 是 C° 的暴露面.

再证 $F^{\triangle}$ 是正常的. 由 $F^{\triangle}$ 的构造，它或者是正常的，或者是空集. 因为 F 是正常暴露面，故存在 C 的支撑超平面 $H(y,1)$，使 $F=C\cap H(y,1)$. 根据前面定义 $F^{\triangle}$ 所作的说明知道，$y\in F^{\triangle}$，即 $F^{\triangle}\neq\varnothing$. 所以 $F^{\triangle}$ 是正常的. ▌

与双配极类似，“△”运算也可对 $F^{\triangle}$ 再运用一次，并表示为 $(F^{\triangle})^{\triangle}=F^{\triangle\triangle}$. 由定理 7.5，如果 F 是 C 的正常暴露面，则 $F^{\triangle\triangle}$ 是 C 的正常暴露面.

定理 7.6 设 C 是 R^n 中包含原点为其内点的紧致凸集，F 是 C 的正常暴露面，则 $F^{\triangle\triangle}=F$. 对于 C° 的正常暴露面 G，相似的结论也成立.

证明 只证明第一部分，第二部分证明相似.

由定义，有

$$F^{\triangle\triangle}=\bigcap_{y\in F^{\triangle}}C\cap H(y,1). \tag{15}$$

因为当且仅当 $H(y,1)$ 是 C 的包含 F 的支撑超平面时，$y\in F^{\triangle}$，故由(15)，$F^{\triangle\triangle}$ 是 C 的包含 F 的所有正常暴露面的交. 但 F 也是 C 的正常暴露面，所以这个交就是面 F. ▌

根据以上证明的结果，可以定义下面的概念.

定义 7.3 设 C 是 R^n 中包含原点为其内点的紧致凸集. 对于 C 的暴露面 F，称由(12)式定义的 $F^{\triangle}$(C°的暴露面)是 F 的共轭面. 对于 C°的暴露面 G，称由(13)式定义的 $G^{\triangle}$(C 的暴露面)是 G 的共轭面.

定理 7.5，定理 7.6 的结果表示，C 和 C°的暴露面可以组成一对相互共轭的面 F 和 G.

§8 凸锥

凸锥是极值问题中的主要研究对象之一.

R^n 中的集合 K 称为以 x_0 为顶点的凸锥，是指它包含每一条从 x_0 出发且通过 K 中的点的半直线，即对于 $\forall x\in K, \lambda>0$，有

$$x_0+\lambda(x-x_0)\in K, \tag{1}$$

如图 21 所示. 特别地，以 $x_0=\theta$ 为顶点的锥的特征是 $\lambda>0$ 时，有

$$\lambda K\subset K. \tag{2}$$

(2) 式表明，K 对正数乘法是封闭的，它是从原点出发且通过 K 中的点的全体半直线的并，可以包含原点，也可以不包含原点.

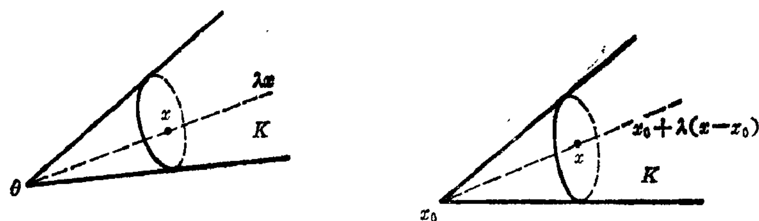


图 21

下面只研究以原点为顶点的凸锥.

1. 凸锥的定义和基本性质

定义 8.1 在 R^n 中，是凸集的锥称为凸锥.

例 1 $R^n, \varnothing, R^n$ 中过原点的闭、开半空间，子空间都是凸

锥.

例 2 R^n 中的两个重要的凸锥是:

1) 非负真锥$\{x=(\xi_1,\cdots,\xi_n)|\xi_1\geqslant 0,\cdots,\xi_n\geqslant 0\}$,

2) 正真锥$\{x=(\xi_1,\cdots,\xi_n)|\xi_1>0,\cdots,\xi_n>0\}$.

定理 8.1 设 K 是 R^n 中的集合, 则 K 是凸锥的充分必要条件是它对于加法和正数乘法封闭.

证明 设 K 是凸锥. 由于 K 是锥,它对正数乘法是封闭的. K 又是凸集, 则对于 $\forall x\in K,\forall y\in K,\lambda=\frac{1}{2}$, 有

$$z=\frac{1}{2}x+\frac{1}{2}y\in K.$$

所以, $x+y=2z\in K$, 即它对加法封闭.

反之, 设 K 对加法和正数乘法封闭. 因为 K 对正数乘法封闭, 故 K 是锥. 对于 $\forall x\in K,\forall y\in K,0<\lambda<1$, 则$(1-\lambda)x\in K$, $\lambda y\in K$. 而由 K 对加法的封闭性,又可以得到

$$(1-\lambda)x+\lambda y\in K,$$

所以K是凸集. 由定义 8.1, K 是凸锥. ▌

推论 8.1.1 设 K 是 R^n 中的子集,则 K 是凸锥的充分必要条件是K包含它的元素的全部正线性组合,即如果 $x_i\in K,\lambda_i>0$, $i=1,\cdots,m$, 则

$$\lambda_1x_1+\cdots+\lambda_mx_m\in K. \tag{3}$$

证明留给读者.

推论 8.1.2 设 S 是 R^n 中的子集, K 是 S 的元素的正线性组合的全体组成的集合. 则 K 是包含 S 的最小凸锥.

证明 由推论 8.1.1, K 是凸锥且 $K\supset S$. 另一方面, 每一个包含 S 的凸锥一定包含 S 的元素的正线性组合的全体,因而一定包含 K. 故 K 是包含 S 的最小凸锥. ▌

推论 8.1.3 设 C 是 R^n 中的凸集,令

$$K=\{\lambda x \mid \lambda>0, x\in C\}=\bigcup_{\lambda>0}\lambda C, \tag{4}$$

则 K 是包含 C 的最小凸锥.

证明 显然.(4)式定义的 K 对正数乘法是封闭的. 设 $\lambda_1 x_1 \in K, \lambda_2 x_2\in K$, 其中 $\lambda_1>0$, $\lambda_2>0$, $x_1\in C$, $x_2\in C$. 因为 C 是凸集, $0<\frac{\lambda_1}{\lambda_1+\lambda_2}<1, 0<\frac{\lambda_2}{\lambda_1+\lambda_2}<1$, 故 $\frac{\lambda_1}{\lambda_1+\lambda_2}x_1+\frac{\lambda_2}{\lambda_1+\lambda_2}x_2 \in C$. 从而

$$\lambda_1 x_1+\lambda_2 x_2=(\lambda_1+\lambda_2)\left(\frac{\lambda_1}{\lambda_1+\lambda_2}x_1+\frac{\lambda_2}{\lambda_1+\lambda_2}x_2\right)\in K.$$

所以 K 对加法封闭, K 是凸锥.

K 的每个点均可看成 C 的点的正线性组合,所以 K 包含在 C 的点的正线性组合的全体所成的集合之中.

设 $x_1,\cdots,x_m\in C$, $\lambda_i>0, i=1,\cdots,m$. 则由于 C 是凸集,知

$$\lambda_1 x_1+\cdots+\lambda_m x_m=(\lambda_1+\cdots+\lambda_m)\left(\frac{\lambda_1}{\lambda_1+\cdots+\lambda_m}x_1+\cdots+\frac{\lambda_m}{\lambda_1+\cdots+\lambda_m}x_m\right)\in K,$$

这表示 C 的点的正线性组合的全体所成的集合包含在 K 中. 所以 K 是 C 的点的正线性组合的全体所成的集合. 由推论 8.1.2, K 是包含 C 的最小凸锥. ▍

定义 8.2 由推论 8.1.2(推论 8.1.3)中的凸锥 K 加上原点所成的集合称为由 $S(C)$ 生成的凸锥或锥包,用 cone$S$ (cone$C$) 表示.

从上面的定义知道, cone$S$ 与包含 S 的最小凸锥一般是有区别的. 如果 $S\neq\varnothing$, cone$S$ 由 S 的点的非负线性组合的全体所成的集合组成. 我们规定, cone$\varnothing=\varnothing$.

定理 8.2 任意个凸锥的交是凸锥.

证明留给读者.

例 3 设 $b_i\in R^n, i\in I$, I 是任意指标集. 则

$$K=\{x\in R^n \mid \langle x,b_i\rangle\leqslant 0, i\in I\} \tag{5}$$

是凸锥，因为它是闭半空间族的交.

(4) 式中的“$\leqslant$”可以用“$\geqslant$”,“$<$”,“$>$”,“$=$”代替，所以联立齐次线性方程和不等式组的解集合是凸锥.

定理 8.3 设 K 是包含原点的凸锥，则

1) $K-K=\{x-y \mid x\in K, y\in K\}=\operatorname{aff}K$ 是包含 K 的最小子空间,

2) $(-K)\cap K$ 是包含在 K 中的最大子空间.

证明 只证明 1), 2) 的证明留给读者.

容易证明 $K-K$ 是子空间. 因为 K 包含原点, 因而 $K\subset K-K$. 假设还有一个子空间 D, 使 $D\supset K$, $D\subset K-K$, 且 $(K-K)\setminus D\neq\varnothing$. 令 $z\in(K-K)\setminus D$, 则 $z\in K-K$, $z\notin D$, 故可设 $z=x-y$, $x\in K$, $y\in K$. 于是 $x\in D$, $y\in D$. 因为 D 是子空间, 故 $x-y\in D$, 矛盾. 所以 $K-K$ 是包含 K 的最小子空间. 因为包含原点的集合的仿射包是子空间, 所以 $K-K=\operatorname{aff}K$. ▌

定理 8.4 设 K_1, K_2 是包含原点的凸锥，则

$$K_1+K_2=\operatorname{co}(K_1\cup K_2).$$

证明留给读者.

2. 共轭凸锥

下面讨论凸锥的对偶理论.

定义 8.3 设 K 是 R^n 中的凸锥，则

$$K^*=\{x^*\in R^n \mid \langle x, x^*\rangle\geqslant 0, x\in K\} \tag{6}$$

称为 K 的共轭锥.

显然, K^* 是凸锥.

例 4 对于例 2 中的非负真锥 K, 可以验证 $K^*=K$.

图 22 是 R^2 中的凸锥及其共轭锥.

定理 8.5 设 K 是 R^n 中的凸锥，则 K^* 是闭凸锥.

证明 设 $x_k^*\in K^*$, $k=1,2,\cdots$, $x_k^*\xrightarrow[k\to\infty]{}x_0^*$. 由定义 8.3, 对 $\forall x\in K$, 有

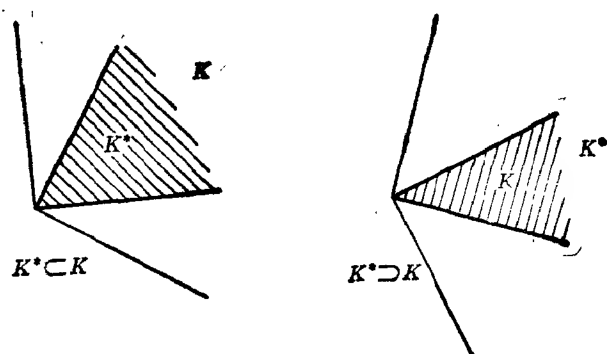


图 22

$$\langle x, x_k^* \rangle \geqslant 0. \tag{7}$$

在(7)式中令 $k\to\infty$时取极限,得

$$\langle x, x_0^* \rangle \geqslant 0,$$

即 $x_0^* \in K^*$,故 K^* 是闭集,因而是闭凸锥. ▌

定理 8.6 设 K 是 R^n 中的凸锥,则

$$K^* = (\mathrm{cl}K)^*.$$

证明留给读者.

定理 8.7 设 K 是 R^n 中的闭凸锥,对于 $\forall x^* \in K^*$,

$$\langle x_0, x^* \rangle \geqslant 0, \tag{8}$$

则 $x_0 \in K$.

证明 假设相反,$x_0 \notin K$. 由定理 5.8,存在 $\bar{x}^*$ 和 $\varepsilon>0$,对于 $\forall x\in K$,满足

$$\langle x_0, \bar{x}^* \rangle \leqslant \langle x, \bar{x}^* \rangle - \varepsilon. \tag{9}$$

下面证明 $\inf\limits_{x\in K}\langle x, \bar{x}^* \rangle = 0$.

因为凸锥包含 x 则一定包含 λx,$\lambda>0$,令 $\lambda\to 0$, 知凸锥一定包含有充分接近于 θ 的点,由 K 是闭集,故 $\theta\in K$,则 $\inf\limits_{x\in K}\langle x, \bar{x}^* \rangle \leqslant 0$.

假设存在$x_1\in K$，使$\langle x_1,\bar{x}^*\rangle<0$. 取$x=\lambda x_1\in K,\lambda>0$，则$\langle x,\bar{x}^*\rangle\xrightarrow[\lambda\to+\infty]{}-\infty$. 但由(9)，$\langle x,\bar{x}^*\rangle$有下界，矛盾. 所以$\inf\limits_{x\in K}\langle x,\bar{x}^*\rangle=0$. 从而对于$\forall x\in K,\langle x,\bar{x}^*\rangle\geqslant0$，故$\bar{x}^*\in K^*$.

在(9)式中令$x=\theta$，得$\langle x_0,\bar{x}^*\rangle\leqslant-\varepsilon$. 这与(8)式的条件矛盾，所以$x_0\in K$. ▌

与配极的情形类似，称$K^{**}=(K^*)^*$为K的双共轭锥.

定理 8.8 设K是R^n中的闭凸锥，则$K^{**}=K$.

证明 由定义8.3知

$$K^{**}=\{x\mid\langle x,x^*\rangle\geqslant0,x^*\in K^*\}. \tag{10}$$

如果$x\in K$，则对于$\forall x^*\in K^*$，$\langle x,x^*\rangle\geqslant0$. 故$x\in K^{**}$，所以$K\subset K^{**}$.

反之，设$x\in K^{**}$. 因为K是闭集，则对于$\forall x^*\in K^*$，有

$$\langle x,x^*\rangle\geqslant0. \tag{11}$$

而由定理8.7知$x\in K$，即$K^{**}\subset K$. 故$K^{**}=K$. ▌

定理8.8中K是闭集的条件是必要的，如果没有K是闭集的条件，只能得到

$$K^{**}=\mathrm{cl}K. \tag{12}$$

定理 8.9 设K_1,K_2是R^n中的凸锥，则K_1+K_2是凸锥且有

$$(K_1+K_2)^*=K_1^*\cap K_2^*, \tag{13}$$

$$(\mathrm{cl}K_1\cap\mathrm{cl}K_2)^*=\mathrm{cl}(K_1^*+K_2^*). \tag{14}$$

证明 由定理3.15，K_1+K_2是凸集. 设$x\in K_1+K_2$，则存在$x_1\in K_1,x_2\in K_2$，使$x=x_1+x_2$. 但在$\lambda>0$时，$\lambda x_1\in K_1,\lambda x_2\in K_2$，故$\lambda x=\lambda x_1+\lambda x_2\in K_1+K_2$，即$K_1+K_2$对正数乘法封闭，所以$K_1+K_2$是锥，于是$K_1+K_2$是凸锥.

由定义8.3，$x^*\in(K_1+K_2)^*$的等价条件是对于$\forall x_1\in K_1$，$\forall x_2\in K_2$，有

$$\langle x_1+x_2,x^*\rangle\geqslant0. \tag{15}$$

因为 x_1, x_2 是彼此独立变化的，在(15)式中分别令 x_1, x_2 趋于 θ，则(15)式又等价于

$$\langle x_1, x^*\rangle \geqslant 0, x_1 \in K_1, \tag{16}$$

$$\langle x_2, x^*\rangle \geqslant 0, x_2 \in K_2 \tag{17}$$

同时成立，即 $x^* \in K_1^*$, $x^* \in K_2^*$. 所以 $x^* \in (K_1 + K_2)^*$ 等价于 $x^* \in K_1^* \cap K_2^*$，故(13)式成立.

从(12)式知

$$K_1^{**} = \mathrm{cl}K_1,\quad K_2^{**} = \mathrm{cl}K_2,$$

则利用(13)式得

$$\begin{aligned}(\mathrm{cl}K_1 \cap \mathrm{cl}K_2)^* &= (K_1^{**} \cap K_2^{**})^* = [(K_1^* + K_2^*)^*]^* \\ &= (K_1^* + K_2^*)^{**} = \mathrm{cl}(K_1^* + K_2^*).\end{aligned} \tag{18}$$

故(14)成立. ▌

定理 8.10 设 K 是 R^n 中的凸锥，且 $\inf\limits_{x \in K}\langle x, x^*\rangle > -\infty$. 则 $x^* \in K^*$. 如果 $x \in \mathrm{int}K$，则对于 $\forall x^* \in K^*$, $x^* \neq \theta$ 时，$\langle x, x^*\rangle > 0$.

证明 在 $\inf\limits_{x \in K}\langle x, x^*\rangle > -\infty$ 时，用定理 8.7 中类似的方法可以证得 $\inf\limits_{x \in K}\langle x, x^*\rangle = 0$，故 $x^* \in K^*$.

如果 $x \in \mathrm{int}K$，则存在 $\varepsilon > 0$，使 $x + \varepsilon B \subset K$，其中 B 是欧氏单位球. 故对于 $x^* \in K^*$ 及 $\forall z \in B$，有

$$\langle x + \varepsilon z, x^*\rangle \geqslant 0, \tag{19}$$

当 $x^* \neq \theta$ 时，利用(19)式得

$$\langle x, x^*\rangle \geqslant \varepsilon \sup_{z \in B}\langle -z, x^*\rangle \geqslant \varepsilon \left\langle \frac{x^*}{|x^*|}, x^* \right\rangle = \varepsilon |x^*| > 0. \quad ▌$$

3. 凸锥的分离定理

由于凸锥是一类具有特殊性质的凸集，所以凸锥分离的概念与凸集分离的概念稍有不同. 我们将要讨论多个凸锥（而不仅是

两个凸锥）分离的重要特点．为了使读者对凸锥分离的概念更易于理解，先从保证凸锥分离的一些条件开始讨论，再定义凸锥的分离．

定理 8.11 设$K_1,\cdots,K_m$是R^n中的凸锥，则$\bigcap_{i=1}^m K_i=\varnothing$的必要条件是存在不全为$\theta$的$x_i^*\in K_i^*,i=1,\cdots,m$，满足

$$x_1^*+\cdots+x_m^*=\theta. \tag{20}$$

证明 考虑元素为$X=(x_1,\cdots,x_m)$的$m\times n$维空间R^{mn}，其中$x_i\in R^n$，$i=1$，$\cdots$，m．显然$X=(x_1,\cdots,x_m)$与$X^*=(x_1^*,\cdots,x_m^*)$的内积可以写成

$$\langle X,X^*\rangle=\langle x_1,x_1^*\rangle+\cdots+\langle x_m,x_m^*\rangle.$$

设

$$\widetilde{K}=\{(x_1,\cdots,x_m)\mid x_1\in K_1,\cdots,x_m\in K_m\},$$

$$\widetilde{P}=\{\overbrace{(x,\cdots,x)}^{m\text{个}}\mid x\in R^n\}.$$

容易验证，$\widetilde{K}$和$\widetilde{P}$都是凸锥．

设$\bigcap_{i=1}^m K_i=\varnothing$，则

$$\widetilde{K}\cap\widetilde{P}=\varnothing.$$

由定理5.5，在R^{mn}中存在不为θ的点$X^*=(x_1^*,\cdots,x_m^*)$，使对于$x\in R^n,x_i\in K_i,i=1,\cdots,m$，满足

$$\langle x,x_1^*\rangle+\cdots+\langle x,x_m^*\rangle\leqslant\langle x_1,x_1^*\rangle+\cdots+\langle x_m,x_m^*\rangle. \tag{21}$$

由(21)式知，$\langle x_i,x_i^*\rangle$在K_i上有下界，$i=1,\cdots,m$．所以由定理8.10得到

$$x_i^*\in K_i^*,i=1,\cdots,m. \tag{22}$$

(21)式也说明$\langle x,x_1^*+\cdots+x_m^*\rangle$在$R^n$中有上界，故

$$x_1^*+\cdots+x_m^*=\theta,$$

这是因为非零线性函数在 R^n 上是无界的. ▌

定理 8.12 设 $K_1,\cdots,K_m$ 是 R^n 中的凸锥,且 $K_1\cap\mathrm{int}K_2\cap\cdots\cap\mathrm{int}K_m\neq\varnothing$. 令 $K=K_1\cap\cdots\cap K_m$. 则

$$K^*=K_1^*+\cdots+K_m^*. \tag{23}$$

证明 根据共轭锥的定义容易证明

$$K^*\supset K_1^*+\cdots+K_m^*. \tag{24}$$

于是只要证明反向的包含关系也成立,就证明了(23)式.

设 $x^*\in K^*,x^*\neq\theta,K_0=\{x\mid\langle x,x^*\rangle<0\}$. 则

$$K_0\cap K=\varnothing.$$

设 $\forall y^*\in K_0^*,y^*\neq\theta,x\in K_0$. 则由 $\langle x,x^*\rangle<0$,而 $\langle x,y^*\rangle\geqslant 0$ 知 x^*,y^* 线性相关,故存在不全为 0 的 α_1,α_2,满足

$$\alpha_1x^*-\alpha_2y^*=\theta.$$

但 $x^*\neq\theta,y^*\neq\theta$,所以 $\alpha_2\neq0$,故 $y^*=\lambda x^*$,其中 $\lambda=\frac{\alpha_1}{\alpha_2}$. 因为 $x\in K_0$,则 $0\leqslant\langle x,y^*\rangle=\lambda\langle x,x^*\rangle$. 注意到 $\langle x,x^*\rangle<0$,于是 $\lambda<0$.

如果 $y^*=\theta$,可以设 $y^*=0\cdot x^*$. 所以

$$K_0^*=\{y^*\mid y^*=\lambda x^*,\lambda\leqslant0\}.$$

因为 $K_0\cap K=\varnothing$,由定理 8.11,存在不全为 θ 的 $y^*\in K_0^*,x_i^*\in K_i^*,i=1,\cdots,m$,满足

$$y^*+x_1^*+\cdots+x_m^*=\theta. \tag{25}$$

根据 K_0^* 的结构,(25)又可改写为

$$-\lambda x^*=x_1^*+\cdots+x_m^*, \tag{26}$$

其中 $\lambda\leqslant0$,下面证明 $\lambda\neq0$。

假设相反,$\lambda=0$。由(26)得

$$x_1^*+\cdots-x_m^*=\theta,$$

且 $x_1^*,\cdots,x_m^*$ 不全为 θ. 这表示至少有两个 x_i^* 不为 θ, 不妨设 $x_1^*\neq\theta,x_2^*\neq\theta$. 由已知条件,取

$$x_0\in K_1\cap\mathrm{int}K_2\cap\cdots\cap\mathrm{int}K_m,$$

则由定理8.10，得

$$\langle x_0, x_2^*\rangle > 0, \langle x_0, x_i^*\rangle \geqslant 0, i \neq 2。$$

于是得出矛盾的结果：

$$0 = \langle x_0, x_1^* + \cdots + x_m^*\rangle = \langle x_0, x_1^*\rangle + \langle x_0, x_2^*\rangle + \cdots + \langle x_0, x_m^*\rangle > 0。$$

所以，$\lambda \neq 0$.

由(26)式得

$$x_* = \left(-\frac{1}{\lambda}\right)x_1^* + \cdots + \left(-\frac{1}{\lambda}\right)x_m^* \in K_1^* + \cdots + K_m^*。\tag{27}$$

所以，$K^* \subset K_1^* + \cdots + K_m^*$. ▌

图 23 是定理 8.12 证明的示意图.

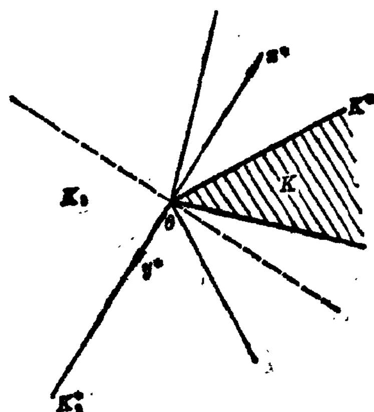


图 23

定理 8.13 设 $K_1, \cdots, K_m$ 是 R^n 中的凸锥，令

$$K = K_1 \cap \cdots \cap K_m。$$

则

$$K^* = K_1^* + \cdots + K_m^* \tag{28}$$

成立或存在不全为 θ 的点 $x_i^* \in K_i^*, i = 1, \cdots, m$，使

$$x_1^* + \cdots + x_m^* = \theta。\tag{29}$$

证明 回到(26)式. 如果对于全部 $x^*\in K^*$, 在 (26) 式中 $\lambda<0$, 则 x^* 可以表示成 $x_i^*\in K_i^*$ 的和, $i=1,\cdots,m$, 这时 (28)式成立. 如果对于某个 $x^*\in K^*$, $\lambda=0$. 则(29)式成立. ▌

从定理 8.11 知, 当 $K_1\cap\cdots\cap K_m=\varnothing$ 时, (29)式成立. 而由定理 8.12 的证明过程又知道, 在(29)式成立时, 交 $K_1\cap\text{int}\,K_2\cap\cdots\cap\text{int}K_m=\varnothing$. 所以我们按下面的方式定义凸锥分离的概念.

定义 8.4 设 $K_1,\cdots,K_m$ 是 R^n 中的凸锥. 如果存在不全为 θ 的 $x_i^*\in K_i^*$, $i=1,\cdots,m$, 使

$$x_1^*+\cdots+x_m^*=\theta, \tag{30}$$

则称凸锥 $K_1,\cdots,K_m$ 可以分离.

定理 8.14 设 K_1,K_2 是 R^n 中的凸锥. 则 K_1,K_2 不能分离的充分必要条件是:

1) $\text{ri}K_1\cap\text{ri}K_2\neq\varnothing$, (31)

2) $\text{Lin}K_1+\text{Lin}K_2=R^n$. (32)

证明 全部证明分为三个部分.

首先证明 K_1,K_2 不能分离的充分必要条件是:

$$\theta\in\text{int}(K_1-K_2). \tag{33}$$

如果 K_1,K_2 可以分离, 则根据定义 8.4, 可以找到 $x_1^*\in K_1^*$, $x_2^*\in K_2^*$, $x_1^*\neq\theta$, 满足 $x_1^*+x_2^*=\theta$. 故对 $\forall x_1\in K_1,\forall x_2\in K_2$, 有

$$\langle x_1-x_2,x_1^*\rangle=\langle x_1,x_1^*\rangle+\langle x_2,x_2^*\rangle\geqslant 0\ . \tag{34}$$

由(34)式立即可以得到, $\theta\notin\text{int}(K_1-K_2)$. 因为如果 $\theta\in\text{int}(K_1-K_2)$, 则当 $\varepsilon>0$ 充分小时, $-\varepsilon x_1^*\in K_1-K_2$, 即存在 $\bar{x}_1\in K_1,\bar{x}_2\in K_2$, 使 $-\varepsilon x_1^*=\bar{x}_1-\bar{x}_2$. 故

$$\langle\bar{x}_1-\bar{x}_2,x_1^*\rangle=-\varepsilon|x_1^*|^2<0, \tag{35}$$

这和(34)式矛盾.

反之, 设 $\theta\notin\text{int}(K_1-K_2)$. 这时 θ 和 $\text{int}\ (K_1-K_2)$ 可以分离, 所以由定理 5.5, 存在 $x^*\neq\theta$, 当 $x_1\in K_1,x_2\in K_2$, 且 $x_1-x_2\in\text{int}(K_1-K_2)$ 时

$$\langle x_1 - x_2, x^* \rangle \geq 0. \tag{36}$$

在(36)式中设 $x_1^* = x^*$, $x_2^* = -x^*$, 分别令 $x_2 \to \theta$, $x_1 \to \theta$, 得 $x_1^* \in K_1^*$, $x_2^* \in K_2^*$, 且 $x_1^* + x_2^* = \theta$. 由定义 8.4 知, K_1, K_2 可以分离.

综上述得到, K_1, K_2 不能分离的充分必要条件是 $\theta \in \text{int}(K_1 - K_2)$.

再证明如果1), 2)至少有一个不满足时, K_1, K_2 可以分离. 从而可知当 K_1, K_2 不能分离时, 1), 2)两个条件都成立.

如果1)不满足, 即 $\text{ri}K_1 \cap \text{ri}K_2 = \varnothing$. 则由定理 5.5, 存在 $x^* \neq \theta$, 对于 $\forall x_1 \in \text{ri}K_1$, $\forall x_2 \in \text{ri}K_2$, 满足

$$\langle x_1, x^* \rangle \geq \langle x_2, x^* \rangle. \tag{37}$$

在(37)式中, 设 $x_1^* = x^*$, $x_2^* = -x^*$, 分别令 $x_1 \to \theta$, $x_2 \to \theta$, 得 $x_1^* \in (\text{ri}K_1)^*$, $x_2^* \in (\text{ri}K_2)^*$, 且 $x_1^* + x_2^* = \theta$. 因为 $\text{cl}K_i = \text{cl}(\text{ri}K_i)$, $K_i^* = (\text{cl}K_i)^* = (\text{ri}K_i)^*$, $i = 1, 2$, 故由定义8.4, K_1, K_2 可以分离.

如果2)不满足, 即 $\text{Lin}K_1 + \text{Lin}K_2 \neq R^n$. 因为对于凸锥 K, $\theta \in \text{cl}K$, 且 $\text{Lin}(\text{cl}K) = \text{Lin}K$, 故 $K \subset \text{Lin}K$. 所以 $K_1 - K_2 \subset \text{Lin}K_1 - \text{Lin}K_2 = \text{Lin}K_1 + \text{Lin}K_2$. 但因2)不满足, $\text{Lin}K_1 + \text{Lin}K_2$ 是 R^n 的真子空间, $(\text{Lin}\,K_1 + \text{Lin}\,K_2)^\perp \neq \varnothing$. 取 $x^* \in (\text{Lin}\,K_1 + \text{Lin}K_2)^\perp$, $x^* \neq \theta$, 则对 $\forall x_1 \in K_1$, $\forall x_2 \in K_2$, 有

$$\langle x_1 - x_2, x^* \rangle = 0.$$

进行和上面类似的推证可知, K_1, K_2 可以分离.

最后证明1), 2)两个条件成立时, $\theta \in \text{int}(K_1 - K_2)$, 从而 K_1, K_2 不可分离. 设 $e_1, \cdots, e_n$ 是 R^n 中的标准正交向量, 因为条件2)成立, $e_k \in \text{Lin}K_1 + \text{Lin}K_2$, $k = 1, \cdots, n$. 故存在 $g_k \in \text{Lin}K_1$, $f_k \in \text{Lin}K_2$, 使 $e_k = g_k - f_k$. 设 $g = g_1 + \cdots + g_n$, $f = f_1 + \cdots + f_n$, 显然 $g \in \text{Lin}K_1$, $f \in \text{Lin}K_2$. 因为条件1)成立, 可取 $x_0 \in \text{ri}K_1 \cap \text{ri}K_2$. 于是可以选择充分小的 $\varepsilon > 0$, 使

$$x_0+\varepsilon\left(g_k-\frac{1}{n+1}g\right)\in K_1, x_0+\varepsilon\left(f_k-\frac{1}{n+1}f\right)\in K_2,$$

$$k=1,\cdots,n,$$

$$x_0-\frac{\varepsilon}{n+1}g\in K_1,\quad x_0-\frac{\varepsilon}{n+1}f\in K_2.$$

设 $e=g-f=e_1+\cdots+e_n$，则

$$y_k=\left[x_0+\varepsilon\left(g_k-\frac{1}{n+1}g\right)\right]-\left[x_0+\varepsilon\left(f_k-\frac{1}{n+1}f\right)\right]$$

$$=\varepsilon\left(e_k-\frac{1}{n+1}e\right)\in K_1-K_2, k=1,\cdots,n, \tag{38}$$

$$y_0=\left(x_0-\frac{\varepsilon}{n+1}g\right)-\left(x_0-\frac{\varepsilon}{n+1}f\right)=\frac{-\varepsilon}{n+1}e\in K_1-K_2 \tag{39}$$

因为 $y_0,y_1,\cdots,y_n$ 仿射无关，故单纯形 $S^n(y_0,y_1,\cdots,y_n)\subset K_1-K_2$。但 $\theta\in\text{int}S^n(y_0,y_1,\cdots,y_n)$，所以 $\theta\in\text{int}(K_1-K_2)$。▌

利用定理8.14，用归纳法不难证明下面的结果.

定理 8.15 设 $K_1,\cdots,K_m$ 是 R^n 中的凸锥，则 $K_1,\cdots,K_m$ 不能分离的充分必要条件是：

1) $\text{ri}K_1\cap\text{ri}K_2\cap\cdots\cap\text{ri}K_m\neq\varnothing$, (40)

2) $\text{Lin}K_1\cap\cdots\cap\text{Lin}K_{j-1}+\text{Lin}K_j=R^n, j=2,\cdots,m.$ (41)

证明留给读者.

4. 锥包的性质

定理 8.16 设 C 是 R^n 中的凸集，如果 $x=\lambda x_1,\lambda>0, x_1\in\text{int}C$，则 $x\in\text{int}(\text{cone}C)$.

证明留给读者.

定理 8.17 设 $C_1,\cdots,C_m$ 是 R^n 中的凸集，$\theta\in C_i, i=1,\cdots,m$. 则

$$\bigcap_{i=1}^{m}\text{cone }C_i=\text{cone}\bigcap_{i=1}^{m}C_i. \tag{42}$$

证明 设 $x\in\text{cone}\bigcap_{i=1}^{m}C_i$，则存在 $x_1\in\bigcap_{i=1}^{m}C_i,\lambda>0$，使 $x=\lambda x_1$。显然，$x\in\text{cone}C_i,i=1,\cdots,m$。故 $x\in\bigcap_{i=1}^{m}\text{cone}C_i$。因而

$$\bigcap_{i=1}^{m}\text{cone } C_i\supset\text{cone}\bigcap_{i=1}^{m}C_i.$$

反之，设 $x\in\bigcap_{i=1}^{m}\text{cone}C_i$。则 $x\in\text{cone}C_i$，$i=1,\cdots,m$，故存在 $x_i\in C_i,\lambda_i>0$，满足 $x=\lambda_i x_i$。故 $\lambda_i^{-1}x\in C_i$，且 $0\leqslant\lambda\leqslant1$时，有

$$\lambda(\lambda_i^{-1}x)=(1-\lambda)\cdot\theta+\lambda(\lambda_i^{-1}x)\in C_i,i=1,\cdots,m \qquad (43)$$

选取 μ 满足：$0<\mu\leqslant\min\limits_{1\leqslant i\leqslant m}\lambda_i^{-1}$，故 $0<\mu\lambda_i\leqslant1,i=1,\cdots,m$。利用(43)式得

$$\mu x=(\mu\lambda_i)(\lambda_i^{-1}x)\in C_i,i=1,\cdots,m,$$

所以 $\mu x\in\bigcap_{i=1}^{m}C_i$。而 $x=\frac{1}{\mu}\cdot\mu x\in\text{cone}\bigcap_{i=1}^{m}C_i$，故

$$\bigcap_{i=1}^{m}\text{cone}C_i\subset\text{cone}\bigcap_{i=1}^{m}C_i,$$

于是(42)式成立。▌

定理 8.18 设 C 是 R^n 中的凸集，则 $x^*\in(\text{cone } C)^*$ 的充分必要条件是对于 $\forall x\in C,\langle x,x^*\rangle\geqslant0$。

证明留给读者。

定理 8.19 设 C 是 R^n 中的凸集，$\theta\in C$，则 $x\in\text{cone } C$ 的充分必要条件是对于充分小的 $\lambda>0,\lambda x\in C$。

证明 设 $x\in\text{cone}C$。则存在 $x_1\in C,\lambda_1>0$，满足 $x=\lambda_1x_1$。当 $\lambda\leqslant\frac{1}{\lambda_1}$，即 $\lambda\lambda_1\leqslant1$ 时，则有

$$\lambda x=\lambda\lambda_1 x_1=(1-\lambda\lambda_1)\cdot\theta+\lambda\lambda_1 x_1\in C.$$

反之，如果对 $\lambda>0$，$\lambda x=x_1\in C$，则 $x=\lambda^{-1}x_1\in\text{cone}C$. ▌

最后介绍一种特殊的凸锥——法线锥，后面将要经常用到.

定义 8.5 设 C 是 R^n 中的凸集，$x^*\in R^n$. 如果 x^* 不与 C 中任何以 a 为端点的线段成锐角，即对 $\forall x\in C, a\in C$，有

$$\langle x-a, x^*\rangle\leqslant 0, \tag{44}$$

则称 x^* 是凸集 C 在点 a 的法线，在点 a 的法线向量的全体所成的集合称为 C 在点 a 的法线锥.

容易证明，法线锥是凸锥.

例 5 设 $\overline{K}(b,\alpha)=\{x\in R^n|\langle x,b\rangle\leqslant\alpha\}$ 是闭半空间. $a\in H(b,\alpha)=\{x\in R^n|\langle x,b\rangle=\alpha\}$. 则对 $\forall x\in\overline{K}(b,\alpha)$，满足 $\langle x-a, b\rangle\leqslant 0$. 所以 b 是 $\overline{K}(b,\alpha)$ 在 a 的法线. 对于 $\lambda>0$，形如 λb 的向量的全体构成 $\overline{K}(b,\alpha)$ 在 a 的法线锥.

图 24 是法线锥的几个例子.

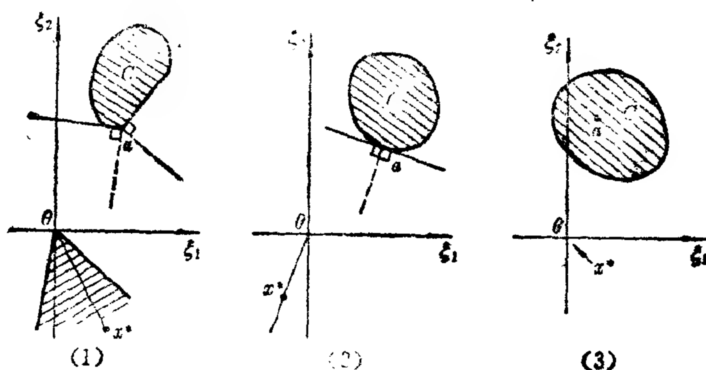


图 24

§ 9 多面体集

多面体集是一类重要的凸集. 它是用有限多个闭半空间的交来定义的. 多面体（多胞形），多面锥都是多面体集的特殊情形.

这一节研究它们的一些性质.

1. 多面体(多胞形)

定义 9.1 设 $x_1, \cdots, x_m$ 是 R^n 中的点, 称由它们生成的凸包 $\mathrm{co}\{x_1, \cdots, x_m\}$ 为多面体, 或多胞形. 如果

$$\dim \mathrm{co}\{x_1, \cdots, x_m\} = k,$$

则称它为 k-多面体.

k 维单纯形 $S^k(x_0, x_1, \cdots, x_k)$ 是 R^n 中 $k+1$ 个仿射无关的点 $x_0, x_1, \cdots, x_k$ 的凸包, 故由定义 9.1 知, 单纯形是多面体的特殊情形. 显然, k-多面体 $\mathrm{co}\{x_1, \cdots, x_m\}$ 中一定存在由 $k+1$ 个仿射无关点组成的子集, 而它的任何一个由 $k+2$ 个点组成的子集一定仿射相关. 应该注意的是, k 维单纯形是 k-多面体, 而 k-多面体不一定是 k 维单纯形, 图 25 就是 R^2 中的两个例子.

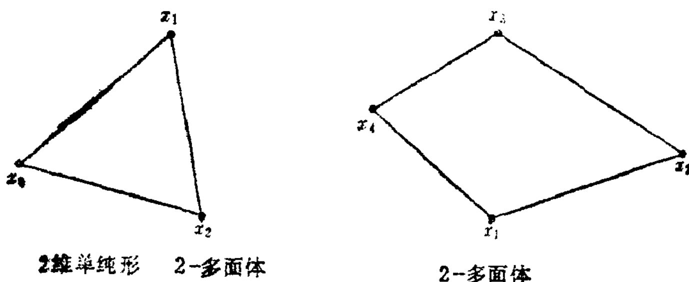


图 25

定理 9.1 设 P 是 R^n 中的非空集, 则下面两个条件等价:

1) P 是多面体,

2) P 是具有有限个极点的紧致凸集.

证明 设 1) 成立, P 是多面体, 即 $P = \mathrm{co}\{x_1, \cdots, x_m\}$. 则由定理 4.13, P 是紧致凸集. 由定理 6.15, $\mathrm{ext}\,P$ 是集合 $\{x_1, \cdots, x_m\}$ 的子集, 故 $\mathrm{ext}\,P$ 是有限点集. 所以条件 2) 成立.

反之, 设 2) 成立, 则由定理 6.14, P 是其极点集 $\mathrm{ext}\,P$ 的凸

包，而 $\text{ext}P$ 是有限点集，故由定义 9.1 知 P 是多面体．所以条件 1）成立．▌

除去仅由一个点组成的多面体外，生成多面体 $P=\text{co}\{x_1, \cdots, x_m\}$ 的集合 $\{x_1, \cdots, x_m\}$ 不是唯一的，因为我们总可以将位于 P 中的另外的点 $x_{m+1}, \cdots$，等加到这个集合中去而不改变 P 的结构．这样，就自然产生一个问题，是否存在生成多面体 P 的唯一最小集合呢？回答是肯定的．下面的定理说明，$\text{ext}P$ 正是生成多面体 P 的唯一最小集合．为了方便，以后也称多面体的极点是它的顶点．

定理 9.2 设 P 是 R^n 中的多面体，则

$$P=\text{co}(\text{ext}P). \tag{1}$$

证明 因为多面体是紧致凸集，故由定理6.14，(1)式成立．▌

定理 9.3 设 P 是 R^n 中的多面体，$\{x_1, \cdots, x_m\}$ 是 P 中的任意有限子集．则下面两个条件等价：

1） $P=\text{co}\{x_1, \cdots, x_m\}$, (2)

2） $\text{ext}P\subset\{x_1, \cdots, x_m\}$. (3)

证明留给读者．

以后将称 $\text{ext}P$ 是多面体 P 的最小表示．

下面研究多面体的面结构．

定理 9.4 设 P 是 R^n 中的多面体，F 是它的正常面，则 F 也是多面体且 $\text{ext}F=F\cap\text{ext}P$.

证明 首先，根据定理 6.8 及定理 9.1 知 P 和 F 都是紧致集．

由定理 6.9，F 的极点集合就是 P 的位于 F 中的极点的集合，即 $\text{ext}F=F\cap\text{ext}P$．而由定理 9.1，$\text{ext}P$ 是有限点集，故 $\text{ext}F$ 也是有限点集．再利用定理 9.1，得到 F 也是多面体．▌

推论 9.4.1 设 P 是 R^n 中的多面体，则 P 的面的个数有限．

证明 由定理 9.1 知 P 的极点数是有限的. 但根据定理 9.4 和定理 9.2, P 的每一个面都是 P 的极点的凸包, 所以面的个数是有限的. ▌

定理 9.5 设 P 是 R^n 中的多面体, 则它的所有的面都是暴露面.

证明 只要对 R^n 中的 n-多面体证明结论成立即可. 下面按 n 进行归纳证明.

对于 $n=0,1,2$, 结论是明显成立的. 设在多面体的维数小于 k, $k\geqslant 3$ 时定理结论也是成立的, 研究 R^k 中的 k-多面体 P.

对于 P 的非正常面, 结论是成立的. 故设 F 是 P 的正常面, 如图 26 所示, 要求 P 的支撑超平面 H_1, 使 $F=H_1\cap P$, 从而 F 也是 P 的暴露面.

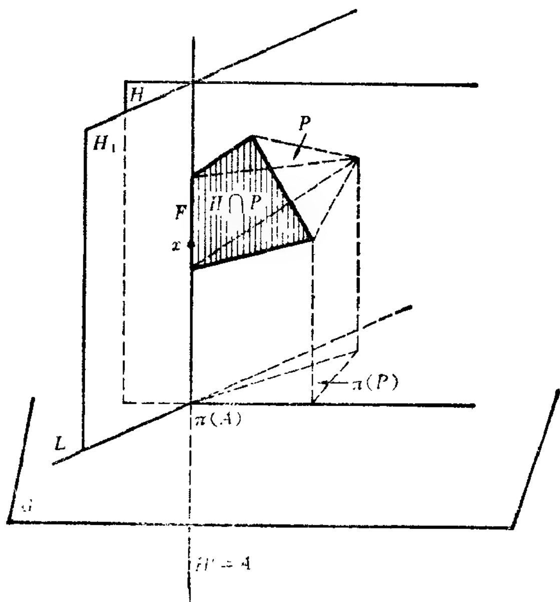


图 26

因为 $F\neq\varnothing$, 故设 $x\in \mathrm{ri}F$, H 是 P 在 x 的正常支撑超平面, 则

$H\cap P$ 是 P 的包含 x 的暴露面. 根据定理 6.11, F 是 P 的包含 x 的最小面,故 $F\subset H\cap P$. 如果 $F=H\cap P$, 则 F 是 P 的暴露面, 这正是要证明的结论.

如果 $F\neq H\cap P$, 根据定理 6.9, F 是 $H\cap P$ 的正常面. 因为 $\dim(H\cap P)<k$, 而 $H\cap P$ 是多面体, 所以由归纳假设知在 $\mathrm{aff}(H\cap P)$ 中存在 $H\cap P$ 的支撑超平面 H', 使 $F=H'\cap(H\cap P)$. 这个超平面也可看成 H 的超平面 A, 满足

$$F=A\cap P. \tag{4}$$

因为 A 是 H 的超平面, H 是 P 的超平面. $\dim P=k$, $k\geqslant 3$, 于是 $\dim A=k-2\geqslant 1$. 设 G 是 R^k 中与 A 正交的二维仿射集. $\pi(P)$ 是 P 在 G 中的正交投影,它是 G 中的 2-多面体. 而 $\pi(A)$ 仅是一点. 可以证明 $\pi(A)$ 是 $\pi(P)$ 的顶点.

事实上,如果设其相反, $\pi(A)$ 不是 $\pi(P)$ 的顶点, 则在 P 中可以找到点 y 和 z, $\pi(y)\neq\pi(z)$, 且对某一个 λ, $0<\lambda<1$, 有

$$\pi(A)=(1-\lambda)\pi(y)+\lambda\pi(z). \tag{5}$$

如果设 $u=(1-\lambda)y+\lambda z$, 则 $u\in P$ 且 $\pi(u)=\pi(A)$. 这表明 $u\in A$, 因为 $\pi^{-1}(\pi(A))=A$, 其中 π^{-1} 表示正交投影变换 π 的逆变换. 由(4)式知 $u\in F$. 因为 F 是 P 的面,故由定义 6.2, $y\in F$, $z\in F$. 但 $F\subset A$, 故对 $\forall v\in F$, $\pi(v)=\pi(A)$. 特别对于 F 中的 y 和 z, $\pi(y)=\pi(z)$, 矛盾. 所以 $\pi(A)$ 是 $\pi(P)$ 的顶点.

因为由归纳假设, 定理在 $k=2$ 时是成立的, 所以在 G 中存在直线 L, 使 $\pi(A)=L\cap\pi(P)$, 则

$$H_1=\mathrm{aff}(A\cup L)$$

是 P 的支撑超平面, 且 $F=H_1\cap P$, 所以 H_1 正是我们要寻求的超平面.

综上所述, 定理的结论对 R^k 中的 k-多面体也是成立的, 所以定理的结论成立. ∎

推论 9.5.1 设 P 是 R^n 中的多面体, 则它的每一个暴露面

都是多面体，且仅有有限个不同的暴露面.

证明留给读者.

定理 9.6 设S是R^n中的k维单纯形，F是S的正常面，则F也是单纯形.

证明 由定理 9.4，面F的顶点就是S的属于F的顶点，因为仿射无关集的任何子集也是仿射无关的，所以F的顶点集合也是仿射无关的．根据定理 9.2，F是其顶点的凸包，故F是单纯形. ▌

定理 9.7 设S是R^n中的k维单纯形，$X \subset \text{ext}S$．令$F=\text{co}X$，则F是S的面，且$\text{ext}F=X$.

证明 设$\text{ext}S=\{x_1,\cdots,x_{k+1}\}$，$X=\{x_1,\cdots,x_m\}$，$k\geq m$．我们将证明，如果$y_0,y_1$是$S$的两个点，对某一个$\lambda$，$0<\lambda<1$，点$y=(1-\lambda)y_0+\lambda y_1\in F$，则$y_0,y_1$也属于$F$，于是由定义 6.2，$F$是$S$的面.

因为$\text{ext}S$是S的最小表示，故对$\forall x\in S$，有

$$x=\lambda_1 x_1+\cdots+\lambda_{k+1}x_{k+1}, \tag{6}$$

其中$\lambda_i\geq 0$，$i=1,\cdots,k+1$，$\sum_{i=1}^{k+1}\lambda_i=1$，且这个表达式是唯一的.

而对于S的属于F的点x，在(6)式中，当$i=m+1,\cdots,k+1$时，$\lambda_i=0$，所以，对y_0,y_1有

$$y_0=\sum_{i=1}^{k+1}\lambda_{0i}x_i,\ \lambda_{0i}\geq 0,\ i=1,\cdots,k+1,\ \sum_{i=1}^{k+1}\lambda_{0i}=1, \tag{7}$$

$$y_1=\sum_{i=1}^{k+1}\lambda_{1i}x_i,\ \lambda_{1i}\geq 0,\ i=1,\cdots,k+1,\ \sum_{i=1}^{k+1}\lambda_{1i}=1. \tag{8}$$

由(7)，(8)两式得

$$y=\sum_{i=1}^{k+1}[(1-\lambda)\lambda_{0i}+\lambda\lambda_{1i}]x_i. \tag{9}$$

另一方面，由 $y\in F$，又有

$$y=\sum_{i=1}^{n}\bar{\lambda}_i x_i. \tag{10}$$

因为(6)式是唯一的，故

$$(1-\lambda)\lambda_{0i}+\lambda\lambda_{1i}=0, i=m+1,\cdots,k+1. \tag{11}$$

所以 $\lambda_{0i}=\lambda_{1i}=0, i=m+1,\cdots,k+1$，即 y_0, y_1 属于 F.

由定理 9.3，知 $\mathrm{ext}F\subset\{x_1,\cdots,x_m\}$. 反向的包含是显然成立的，所以 $\mathrm{ext}F=X$. ▌

下面的结果是与多面体的外表示有关的，它们说明多面体是可以表示成有限个闭半空间的交的有界集.

定理 9.8 设凸集

$$C=\{x\mid\langle x,x_i^*\rangle\leqslant\alpha_i, i\in I\}, \tag{12}$$

其中 $I=\{1,\cdots,m\}, x_i^*\in R^n, \alpha_i\in R, i\in I, x_0\in C$. 令

$$I(x_0)=\{i\mid\langle x_0,x_i^*\rangle=\alpha_i, i\in I\}.$$

则 $x_0\in\mathrm{ext}C$ 的充分必要条件是集合 $\{x_i^*\mid i\in I(x_0)\}$ 中包含 n 个线性无关的元素.

在 R^3 中，定理 9.8 中的条件的几何解释是：C 的极点 x_0 是界定 C 的三个不通过同一直线的平面的交点.

证明 必要性. 设 x_0 是 C 的极点，而$\{x_i^*\mid i\in I(x_0)\}$中线性无关的元素个数小于 n. 于是关于 x 的线性方程组

$$\langle x,x_i^*\rangle=0, i\in I(x_0),$$

有非零解 $\bar{x}$. 则由 $I(x_0)$的定义及 $x_0\in C$，当 $\varepsilon>0$ 充分小时，有

$$\langle x_0\pm\varepsilon\bar{x},x_i^*\rangle=\alpha_i, i\in I(x_0), \tag{13}$$

$$\langle x_0\pm\varepsilon\bar{x},x_i^*\rangle<\alpha_i, i\in I\setminus I(x_0). \tag{14}$$

所以得到

$$x_0+\varepsilon\bar{x}\in C, x_0-\varepsilon\bar{x}\in C.$$

$$x_0=\frac{1}{2}(x_0+\varepsilon\bar{x})+\frac{1}{2}(x_0-\varepsilon\bar{x})\in C.$$

于是 x_0 不是C的极点，与假设矛盾．所以 $\{x_i^*|i\in I(x_0)\}$ 包含 n 个线性无关的元素．

充分性　设 $x_0\in C,\{x_i^*|i\in I(x_0)\}$ 中包含 n 个线性无关的元素．将定义 C 的不等式组改写为

$$\langle x_0,x_i^*\rangle=\alpha_i,i\in I_0, \tag{15}$$

$$\langle x_0,x_i^*\rangle\leqslant\alpha_i,i\in I\backslash I_0, \tag{16}$$

其中 I_0 是 $I(x_0)$ 中 n 个线性无关元素 x_i^* 对应的指标集．假设 x_0 不是C的极点，即存在一个 $\lambda,0<\lambda<1,x_1\in C,x_2\in C,x_1\neq x_2$，满足

$$x_0=(1-\lambda)x_1+\lambda x_2. \tag{17}$$

根据(15)，(16)式，有

$$\langle x_k,x_i^*\rangle\leqslant\alpha_i,i\in I,k=1,2. \tag{18}$$

而由(15)，(18)式，只有

$$\langle x_k,x_i^*\rangle=\alpha_i,i\in I_0,k=1,2 \tag{19}$$

时(17)式才成立．但线性方程组

$$\langle x,x_i^*\rangle=\alpha_i,i\in I_0, \tag{20}$$

中方程的个数为 $n,\{x_i^*|i\in I_0\}$ 线性无关，它具有唯一解．这与(19)式是矛盾的．所以 x_0 是C的极点．▌

定理 9.9　设C是R^n中的有界集，由

$$C=\{x|\langle x,x_i^*\rangle\leqslant\alpha_i,i\in I\} \tag{21}$$

定义，其中 I 是有限指标集，$x_i^*\in R^n,\alpha_i\in R,i\in I$，则 C 是多面体．

证明　因为C有界，且是闭半空间的交，故它是紧致凸集．由定理 6.14，它是其极点的凸包．再由定理 9.8，C 的每一个极点是形如具有 n 个变量的 n 个线性方程的方程组(15)式的唯一解，所以这个极点由指标集 I_0 单值地确定．因为 I 是有限的，从中选出具有 n 个元素的子集的选法也是有限的，所以极点个数有限．根据定理 9.1 的条件知C是多面体．▌

定理 9.9 说明有限个闭半空间的交形成的有界集或有限个线性不等式组成的不等式组的解形成的有界集是多面体.

设 $x_1,\cdots,x_m$ 是 R^n 中的点，$C=\mathrm{co}\{x_1,\cdots,x_m\}$ 是多面体，令函数

$$\delta^*(x^*|C)=\sup\{\langle x,x^*\rangle|x\in C\}=\max_{1\le i\le m}\langle x_i,x^*\rangle, \quad (22)$$

及 $I(x^*)=\{i|\langle x_i,x^*\rangle=\delta^*(x^*|C),i=1,\cdots,m\}$. 则由(22)知 $H_{x^*}=\{x|\langle x,x^*\rangle=\delta^*(x^*|C)\}$ 是 C 的支撑超平面.

显然,多面体具有无穷多个支撑超平面. 但是,如果从构造多面体出发,则并不需要这无穷多个支撑超平面,而只需要其中某些具有特殊的性质的支撑超平面就够了. 为了研究这些支撑超平面，我们给出下面的定义.

定义 9.2 设 $C=\mathrm{co}\{x_1,\cdots,x_m\}$, $x_1,\cdots,x_m$ 是 R^n 中的点，$j\in I(x^*)$. 如果在 $i\in I(x^*)$ 时 x_i-x_j 有 $n-1$ 个线性无关,则称 H_{x^*} 是 C 的极支撑超平面.

因为当 $i,j\in I(x^*)$ 时, x_i,x_j 都在 H_{x^*} 上,所以对于极支撑超平面 H_{x^*},由 $x_i,x_j(i=1,\cdots,n-1)$ 生成的凸包是以 x_i,x_j 为顶点的 $n-1$ 维单纯形. 在图 27 中,四面体 $OABC$ 由向量 θ,x_1,x_2,x_3 作凸包而得,分别过它的四个面($\triangle ABC,\triangle ABO,\triangle ACO,\triangle BCO$)的支撑超平面都是极支撑超平面，而仅过它的任何一条棱（例如 $\overline{x_1x_2}$）或仅过它的任何一个顶点（例如 x_1)的支撑超平面都不是极支撑超平面.

定理 9.10 设 C 是 R^n 中的多面体, $\mathrm{int}C\neq\varnothing$, $x_0\notin C$, 则可以用极支撑超平面分离 x_0 和 C,即可以找到一个定义极支撑超平面的 x^*,满足对于 $\forall x\in C$,有

$$\langle x,x^*\rangle<\langle x_0,x^*\rangle. \quad (23)$$

证明 证明的思路是,如果分离 x_0 和 C 的支撑超平面 H_{x^*} 不是极支撑超平面,则利用 x_0 将 x^* 修正为某一个向量 $x^*+a_p\bar{x}^*$,使

指标为 $i\in I(x^*+a, \bar{x}^*)$ 的向量 x_i-x_j 中线性无关的向量数比指标为 $i\in I(x^*)$ 的向量 x_i-x_j 中线性无关的向量数多一个，因为多面体的顶点数是有限的，从而可以在有限步骤内作出极支撑超平面来.

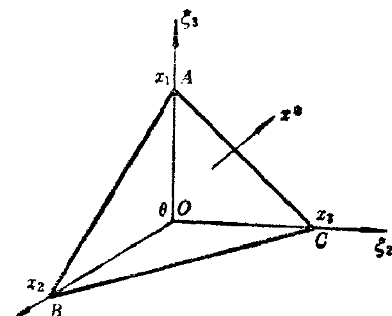


图 27

不失一般性，设 $\theta\in\operatorname{int}C$. 因为多面体 C 是紧致凸集，则由定理 5.8，存在 $x^*\in R^n$ 及 $\varepsilon>0$，对 $\forall x\in C$，有

$$\langle x, x^*\rangle\leqslant\langle x_0, x^*\rangle-\varepsilon, \tag{24}$$

显然

$$\delta^*(x^*|C)<\langle x_0, x^*\rangle. \tag{25}$$

如果 x^* 确定的支撑超平面 H_{x^*} 是极支撑超平面，定理得证. 如果 x^* 不确定极支撑超平面，由定义 9.2，向量集 $\{x_i-x_j\}$ 中 $(i\in I(x^*)$，j 是 $I(x^*)$ 中一个固定的元素) 线性无关的向量不超过 $n-2$ 个，而向量集 $\{x_0-x_j, x_i-x_j, i、j\in I(x^*)\}$ 中线性无关的向量不超过 $n-1$ 个，所以存在 $\bar{x}^*$，满足

$$\langle x_i-x_j, \bar{x}^*\rangle=0, \tag{26}$$

$$\langle x_0-x_j, \bar{x}^*\rangle=0, \tag{27}$$

$$\langle x_0, \bar{x}^*\rangle\leqslant 0. \tag{28}$$

容易证明，至少存在一个 $q\in\{1,\cdots,m\}$，满足

$$\langle x_q, \bar{x}^*\rangle>0,$$

这是因为 $\theta\in\operatorname{int}C$，于是 $0<\delta^*(\bar{x}^*|C)=\max\limits_{1\leqslant i\leqslant m}\langle x_i, \bar{x}^*\rangle$. 由 (26)—(28) 式，当 $i\in I(x^*)$ 时，$\langle x_i, \bar{x}^*\rangle\leqslant 0$，所以 $q\notin I(x^*)$.

当 $i\notin I(x^*)$ 时，用下面的等式确定 $a_i\in R$：

$$\langle x_i, x^* + \alpha_i \bar{x}^* \rangle = \delta^*(x^* | C) + \alpha_i \langle x_0, \bar{x}^* \rangle, \tag{29}$$

故

$$\alpha_i = \frac{\delta^*(x^* | C) - \langle x_i, x^* \rangle}{\langle x_i, \bar{x}^* \rangle - \langle x_0, \bar{x}^* \rangle}. \tag{30}$$

由(26)—(28)式及 $I(x^*)$ 的定义,(30)式的分子总是正的,而 $\langle x_0, \bar{x}^* \rangle < 0$,对于 q, $\langle x_q, \bar{x}^* \rangle > 0$,所以 $\alpha_q > 0$. 在 $i \notin I(x^*)$ 的指标之中,取正的 α_i 中的最小者,记为 α_p,其对应的点为 x_p, $p \in \{1, \cdots, m\}$. 在 $i \notin I(x^*)$ 时,$\langle x_i, x^* \rangle < \delta^*(x^* | C)$,

故

$$\langle x_i, x^* + \alpha_p \bar{x}^* \rangle \leqslant \delta^*(x^* | C) + \alpha_p \langle x_0, \bar{x}^* \rangle, \tag{31}$$

且

$$\langle x_p, x^* + \alpha_p \bar{x}^* \rangle = \delta^*(x^* | C) + \alpha_p \langle x_0, \bar{x}^* \rangle. \tag{32}$$

当 $i \in I(x^*)$ 时,由(26)—(28)式及 $I(x^*)$ 的定义,对 $\forall \alpha \in R$,均有

$$\begin{aligned} \langle x_i, x^* + \alpha \bar{x}^* \rangle &= \langle x_i, x^* \rangle + \alpha \langle x_i, \bar{x}^* \rangle \\ &= \delta^*(x^* | C) + \alpha \langle x_0, \bar{x}^* \rangle. \end{aligned} \tag{33}$$

结合(31)—(33)式,得

$$\delta^*(x^* + \alpha_p \bar{x}^* | C) = \delta^*(x^* | C) + \alpha_p \langle x_0, \bar{x}^* \rangle, \tag{34}$$

$$I(x^* + \alpha_p \bar{x}^*) \supset I(x^*) \cup \{p\}. \tag{35}$$

因为 $\alpha_p > 0$,由(30)式得

$$\langle x_p - x_j, \bar{x}^* \rangle = \langle x_p, \bar{x}^* \rangle - \langle x_0, \bar{x}^* \rangle > 0,$$

所以向量集 $\{x_p - x_j, x_i - x_j, i \in I(x^*)\}$ 线性无关. 于是在向量集 $\{x_i - x_j, i \in I(x^* + \alpha_p \bar{x}^*)\}$ 中线性无关的向量个数至少比向量集 $\{x_i - x_j, i \in I(x^*)\}$ 中线性无关的向量个数多一个.

由(25),(34)两式知,对于 $\forall x \in C$,有

$$\begin{aligned} \langle x, x^* + \alpha_p \bar{x}^* \rangle &\leqslant \delta^*(x^* + \alpha_p \bar{x}^* | C) \\ &= \delta^*(x^* | C) + \alpha_p \langle x_0, \bar{x}^* \rangle \\ &< \langle x_0, x^* \rangle + \alpha_p \langle x_0, \bar{x}^* \rangle \end{aligned}$$

$$=\langle x_0, x^* + \alpha_p \bar{x}^* \rangle,$$

即 $x^* + \alpha_p \bar{x}^*$也满足类似于(24)的不等式，x_0和C可以分离.

如果在向量集$\{x_i - x_j, i \in I(x^* + \alpha_p \bar{x}^*)\}$中线性无关的向量个数等于 $n-1$，则 $x^* + \alpha_p \bar{x}^*$即为所求的向量．反之，则以 $x^* + \alpha_p \bar{x}^*$重复上述过程．因为$\{1, \cdots, m\}$有限，$\text{int}C \neq \varnothing$，就一定可以在有限步骤内作出分离 x_0 和C的极支撑超平面．▌

定理 9.11 设C是R^n中的多面体，则C可以用有限的线性不等式组确定.

证明 如果 $\text{int}C \neq \varnothing$，则由定理 9.10 及定义 9.2 知，$C$的极支撑超平面的个数有限．设 $x_k^*, k=1, \cdots, l$ 是极支撑超平面对应的向量，于是不等式组

$$\langle x, x_k^* \rangle \leq \delta^*(x_k^* | C), k=1, \cdots, l, \tag{36}$$

定义了C．事实上，$\forall x \in C$ 均满足 (36)式，而对 $x \notin C$ 则至少破坏了(36)式中的一个不等式.

如果 $\text{int}C = \varnothing$，则取 $x_0 \in C$，而在子空间$\text{Lin}C$中集合$C - x_0$应包含内点，故可以在 $\text{Lin}C$中用有限的不等式组确定 $C - x_0$．但子空间可以用有限的线性方程组定义，所以定理的结论仍成立．▌

由定义 9.1 及定理 9.9、定理 9.11 知道，多面体既可定义为有限个点的凸包，也可以定义为满足有限个线性不等式的不等式组的有界集.

2．多面锥

定义 9.3 设K是 R^n 中的凸锥，如果在 R^n 中存在有限个点 $x_1, \cdots, x_m$，当且仅当 $x \in K$时，有

$$x = \sum_{i=1}^{m} \lambda_i x_i, \lambda_i \geq 0, i=1, \cdots, m, \tag{37}$$

则称K是多面锥.

定理 9.12 设K是定义 9.3 中的多面锥，则K可以用有限的

齐次线性不等式组定义，即

$$K=\{x|\langle x,x_k^*\rangle\geqslant 0,k=1,\cdots,l\}. \tag{38}$$

证明 研究

$$C=\{x|x=\sum_{i=1}^{m}\lambda_i x_i,\ \lambda_i\geqslant 0,\ i=1,\cdots,m,\sum_{i=1}^{m}\lambda_i\leqslant 1\}.$$

不难看出，C是由$\theta,x_1,\cdots,x_m$生成的多面体. 由定理 9.11，C可以由有限的线性不等式组

$$\langle x,x_k^*\rangle\geqslant\alpha_k,k=1,\cdots,l_1 \tag{39}$$

定义. 因为$\theta\in C$，故由(39)式知$\alpha_k\leqslant 0,k=1,\cdots,l_1$. 容易看出，至少有一个$\alpha_k=0$，故可以设$k=1,\cdots,l\leqslant l_1$时$\alpha_k=0$. 即在$1\leqslant k\leqslant l$时，(39)式变成(38)式中的不等式.

设x满足(37)式，在$\alpha>0$时，令$\alpha x=\sum_{i=1}^{m}\alpha\lambda_i x_i$，则当$\alpha$充分小，使$\sum_{i=1}^{m}\alpha\lambda_i\leqslant 1$时，$\alpha x\in C$，故$\alpha x$满足(39)式，因而满足(38)式中的齐次线性不等式组，但$\alpha>0$，所以x满足(38)式的齐次线性不等式组.

反之，设x满足(38)式，则在$\alpha>0$充分小时，αx满足(39)式，即$\alpha x\in C$. 这表明

$$\alpha x=\sum_{i=1}^{m}\lambda_i x_i,\ \sum_{i=1}^{m}\lambda_i\leqslant 1,\lambda_i\geqslant 0,i=1,\cdots,m.$$

所以，$x=\sum_{i=1}^{m}\frac{\lambda_i}{\alpha}x_i$，故$x$也满足(37)式. ▌

定理 9.13 多面锥是闭凸锥.

证明留给读者.

定理 9.14 设R^n中的凸锥K由有限的齐次线性不等式组

$$\langle x, x_k^* \rangle \geqslant 0, k=1,\cdots,l \tag{40}$$

定义，则K的共轭锥K^*是多面锥，且

$$K^* = \{x^* | x^* = \sum_{k=1}^{l} \gamma_k x_k^*, \gamma_k \geqslant 0, k=1,\cdots,l\}.$$

证明 研究多面锥

$$\tilde{K} = \{x^* | x^* = \sum_{k=1}^{l} \gamma_k x_k^*, \gamma_k \geqslant 0, k=1,\cdots,l\}.$$

由定理 9.13，$\tilde{K}$是闭凸锥．由定义 8.3，如果x满足

$$\langle x, \sum_{k=1}^{l} \gamma_k x_k^* \rangle \geqslant 0, \gamma_k \geqslant 0, k=1,\cdots,l, \tag{41}$$

则$x \in (\tilde{K})^*$．而仅当

$$\langle x, x_k^* \rangle \geqslant 0, k=1, \cdots, l,$$

即$x \in K$时，(41)式成立．所以$K = (\tilde{K})^*$．因为$\tilde{K}$是闭集，于是有

$$K^* = (\tilde{K})^{**} = \tilde{K}. \blacksquare$$

定理 9.15 由齐次线性不等式组定义的凸锥，是多面锥．

证明 设R^n中的凸锥K由(40)式定义．由定理 9.14，共轭锥K^*是多面锥．于是根据定理 9.12，存在$x_1, \cdots, x_m$，当且仅当$x^* \in K^*$时，有

$$\langle x_i, x^* \rangle \geqslant 0, i=1,\cdots,m.$$

因为K是闭集，故$K = K^{**}$．再利用定理 9.14，得到$x \in K$时，有

$$x = \sum_{i=1}^{m} \lambda_i x_i, \lambda_i \geqslant 0, i=1,\cdots,m.$$

由定义 9.3，K是多面锥．

从定理 9.12 和定理 9.15 知道，可以用两种方式定义多面锥．一种方式是定义 9.3，另一种方式是通过有限的齐次线性不

等式组.

定理 9.16 设 K_1、K_2 是 R^n 中的多面锥，则 K_1+K_2，$K_1\cap K_2$ 也是多面锥.

证明留给读者.

定理 9.17 设 $K_1,\cdots,K_m$ 是 R^n 中的多面锥，则

$$(K_1\cap\cdots\cap K_m)^*=K_1^*+\cdots+K_m^*. \tag{42}$$

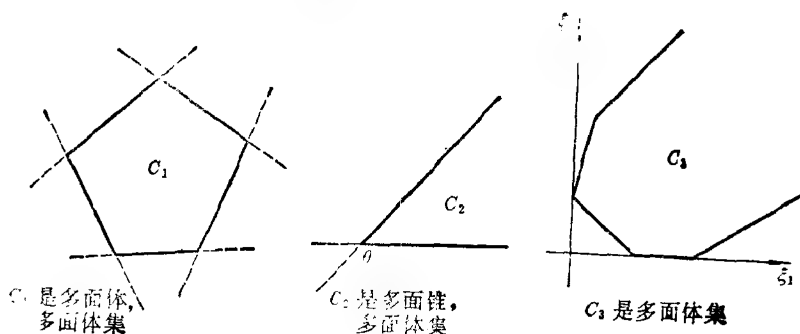
证明 由定理 8.9 知

$$(K_1\cap\cdots\cap K_m)^*=\mathrm{cl}(K_1^*+\cdots+K_m^*).$$

但由定理 9.14 知，K_i^* 是多面锥，$i=1,\cdots,m$，从而 $K_1^*+\cdots+K_m^*$ 也是多面锥. 最后由定理 9.13，$K_1^*+\cdots+K_m^*$ 是闭集，所以(42)式成立. ▌

3. 多面体集

从上面两个小节已经知道，多面体是由有限个闭半空间的交构成的有界集，而多面锥是由有限个过原点的闭半空间的交构成的凸锥，是无界集. 现在将问题一般化，讨论由有限个闭半空间的交构成的集合——多面体集. 我们将会看到，有界的多面体集就是多面体，是凸锥的多面体集就是多面锥.



C_1 是多面体，多面体集　　C_2 是多面锥，多面体集　　C_3 是多面体集

图 28

定义 9.4 设 $C\subset R^n$，如果 C 可以表示成有限个闭半空间的

交，即

$$C=\bigcap_{i=1}^{m}\overline{K}(x_i^*,\alpha_i),$$

其中 $\overline{K}(x_i^*,\alpha_i)=\{x|\langle x,x_i^*\rangle\geqslant\alpha_i\}$，$x_i^*\in R^n$，$\alpha_i\in R$，$i=1,\cdots,m$，则称 C 是多面体集.

例 1 在图 28 中，C_1、C_2、C_3 都是多面体集，C_1 是多面体但不是多面锥.C_2 是多面锥但不是多面体，C_3 不是多面体也不是多面锥.

例 2 R^n 中的每一个超平面 H 是边界为 H 的两个闭半空间的交. 所以 H 是多面体集.

例 3 R^n 中的每一个不与 R^n 重合的仿射集 A 是 R^n 中的有限个超平面的交，而由例 2 的结果，故 A 是多面体集.

由定义 9.4 知，多面体集是闭凸集，有限个多面体集的交仍是多面体集. 多面体集的平移也是多面体集.

定理 9.18 R^n 中的多面体集是多面体与多面锥的和. 反之，多面体与多面锥的和是多面体集.

证明 引入辅助坐标 x^0，则不等式组

$$\langle x,x_i^*\rangle-x^0\alpha_i\geqslant 0,\ i=1,\cdots,l,\ x^0\geqslant 0. \tag{43}$$

是 R^{n+1} 中的齐次线性不等式组. 由定理 9.15，它定义一个多面锥，其元素可以表示为

$$\begin{pmatrix}x^0\\x\end{pmatrix}=\sum_{j=1}^{m}\lambda_j\begin{pmatrix}x_j^0\\x_j\end{pmatrix},\ \lambda_j\geqslant 0,\ j=1,\cdots,m \tag{44}$$

其中 $\begin{pmatrix}x^0\\x\end{pmatrix}$ 表示 R^{n+1} 中分量为 $x^0,x^1,\cdots,x^n$ 的 $n+1$ 维向量. 对于任意的 $\lambda_j\geqslant 0$，(44) 式右边应满足 (43)，特别地，当 $\lambda_j=1$，$i\neq j$ 时 $\lambda_i=0$，便推出 $x_j^0\geqslant 0$.

设

$$I^0=\{j\mid x_j^0=0,\ j=1,\cdots,m\},$$
$$I^+=\{j\mid x_j^0>0,\ j=1,\cdots,m\}.$$

在 $j\in I^+$ 时，令 $y_j=\frac{x_j}{x_j^0}$，$\gamma_j=\lambda_j x_j^0$，则(44)式改写成

$$x^0=\sum_{j\in I^+}\lambda_j x_j^0=\sum_{j\in I^+}\gamma_j,\ \gamma_j\geqslant 0,\ j\in I^+, \tag{45}$$

$$\begin{aligned}x&=\sum_{j\in I^0}\lambda_j x_j+\sum_{j\in I^+}\lambda_j x_j^0\cdot\frac{x_j}{x_j^0}\\&=\sum_{j\in I^0}\lambda_j x_j+\sum_{j\in I^+}\gamma_j y_j,\ \gamma_j\geqslant 0,\ j\in I^+.\end{aligned} \tag{46}$$

所以不等式组(43)式的解可以表示成(45)，(46)式的形式．但只要在(43)式中令 $x^0=1$，就得到定义多面体集的不等式组．所以这个不等式组的每一个解可以表示成

$$x=\sum_{j\in I^0}\lambda_j x_j+\sum_{j\in I^+}\gamma_j y_j,$$
$$\lambda_j\geqslant 0,\ j\in I^0,\ \gamma_j\geqslant 0,\ j\in I^+,\ \sum_{j\in I^+}\gamma_j=1. \tag{47}$$

(47)式右边第一部分和式是多面锥的点，第二部分和式是多面体的点，所以多面体集是多面体与多面锥的和．

反之，从(47)式出发，经(46)，(45)式到(43)式，容易证明，多面体与多面锥的和是多面体集．▌

下面讨论 R^n 中非空多面体集 C 的面结构．如果 C 是 R^n 中的仿射子集，则它的面只有 $\varnothing$ 和 C．如果 C 是 R^n 中的 k 维多面体集但不是仿射集，则 C 与 R^k 中一个不等于 R^k、维数为 k 的多面体集 C' 同构．所以在研究多面体集的面的性质时，只要研究 R^n 中不等于 R^n、维数为 n 的多面体集就行了．

定理 9.19 设 C 是 R^n 中的 n 维多面体集，$C\neq R^n$，且

$$C=\bigcap_{i=1}^{m}\overline{K}(x_i^*,\alpha_i), \tag{48}$$

其中 $\overline{K}(x_i^*,\alpha_i)$是定义 9.4 中的闭半空间，$i=1,\cdots,m$. 则下面的结果成立：

1) $\mathrm{bd}\,C=\bigcup_{i=1}^{m} H(x_i^*,\ \alpha_i)\cap C$,

2) C的每一个超面的形式为$H(x_i^*,\alpha_i)\cap C$,

其中 $\mathrm{bd}\,C$ 是 C 的边界，$H(x_i^*,\alpha_i)$是$\overline{K}(x_i^*,\alpha_i)$的边界.

证明 1) 因为

$$\mathrm{int}C=\mathrm{int}\bigcap_{i=1}^{m}\overline{K}(x_i^*,\alpha_i)=\bigcap_{i=1}^{m}\mathrm{int}\overline{K}(x_i^*,\alpha_i)$$

$$=\bigcap_{i=1}^{m}\overline{K}(x_i^*,\alpha_i)\backslash H(x_i^*,\alpha_i),$$

所以 1) 的结果成立.

2) 设F是C的超面，$x\in\mathrm{ri}F$，则由定理 6.11，F是C的包含x的最小面. 由 1)，存在一个j，使

$$x\in H(x_j^*,\alpha_j)\cap C,$$

从而有

$$F\subset H(x_j^*,\alpha_j)\cap C.$$

因为F与$H(x_j^*,\alpha_j)\cap C$ 都是C的面，且 $\dim F=n-1$，故由推论 6.10.2 得到

$$F=H(x_j^*,\alpha_j)\cap C.\ \blacksquare$$

定理 9.20 设C是R^n中的多面体集，F是C的正常面. 则在C中存在包含F的超面G.

证明 不妨设$\dim C=n$，C由(48)式定义，$x\in\mathrm{ri}F$. 由定理 9.19 的 1)，可以找到一个j，使

$$x\in H(x_j^*,\alpha_j)\cap C.$$

由定理 6.11，可设F是包含x的最小面，$H(x_j^*,\alpha_j)\cap C$是包含x的超面. 令

$$G=H(x_j^*,\alpha_j)\cap C,$$

则G是包含F的超面. ▍

推论 9.20.1 设C是 R^n 中的多面体集, 则C的每一个面都是多面体集.

证明 因为只有$\varnothing$和C本身是C的非正常面，所以仅对C的正常面证明推论的结论. 由定理 9.20，C的每一个正常面都是它的某一个超面的面. 而由定理 9.19 的 2)，C的超面也是多面体集. 于是按维数进行归纳证明就可以得到要证明的结论，其细节留给读者. ▍

推论 9.20.2 R^n 中的每一个多面体集C的面的个数有限.

证明留给读者.

推论 9.20.3 设C是 R^n中的n维多面体集; F_j, F_k是它的两个面，满足

$$F_j\subset F_k,\ \dim F_j=j,\ \dim F_k=k,$$

其中 $0\leqslant j<j+1$, $k-1<k\leqslant n$. 则C中存在面 F_{j+1}, $\cdots$, F_{k-1}, 满足

$$F_j\subset F_{j+1}\subset\cdots\subset F_{k-1}\subset F_k,$$

$$\dim F_i=i,\ i=j+1,\cdots,k-1.$$

证明 根据定理 6.9, F_j 是 F_k 的正常面, 由推论 9.20.1, F_k 是多面体集. 所以, 由定理 9.20, 在 F_k 中存在超面 F_{k-1}, 满足 $F_j\subset F_{k-1}$. 如果 $j=k-2$, 结论得证. 如果 $j<k-2$, 则以 F_{k-1} 代替 F_k, 重复上述过程. 如此继续，则可以得到推论结果要求的面. ▍

最后，讨论多面体和多面体集的对偶关系.

定理 9.21 设 $x_1,\cdots,x_m$是 R^n 中两两互不相同的点，$m\geqslant 1$, 令

$$C=\mathrm{co}\{x_1,\cdots,x_m\},\ D=\bigcap_{i=1}^{m}\bar K(x_i,1),$$

其中 $\overline{K}(x_i,1)=\{x\mid\langle x,x_i\rangle\leqslant 1\}$. 则下面的结论成立：

1) $C^\circ=D$,

2) $D^\circ = \mathrm{co}\{\theta,x_1,\cdots,x_m\}$,

3) 当且仅当 $\theta\in C$ 时，C 和 D 两两配极，

4) 当且仅当 $\theta\in\mathrm{int}\,C$ 时，C 和 D 两两配极，D 有界.

证明 1) 由 §7 的(2)式推出

$$\{x_1,\cdots,x_m\}^\circ=D. \tag{49}$$

同样，由 §7 的(2)式知，当且仅当 $y\in M^\circ$ 时，$M\subset\overline{K}(y,1)$，而当且仅当 $\mathrm{co}M\subset\overline{K}(y,1)$ 时，$M\subset\overline{K}(y,1)$（注意 M 与 $\overline{K}(y,1)$ 的含义见 §7 的(2)式），所以有

$$\{x_1,\cdots,x_m\}^\circ=(\mathrm{co}\{x_1,\cdots,x_m\})^\circ. \tag{50}$$

由(49)，(50)式得 $C^\circ=D$.

2) 利用 1) 式、定理 7.3 及 C 是紧致凸集，得

$$D^\circ=C^{\circ\circ}=\mathrm{cl\,co}\,\{\theta,x_1,\cdots,x_m\}$$
$$=\mathrm{co}\{\theta,x_1,\cdots,x_m\}.$$

即 2) 成立.

3) 直接由 1) 及 2) 得到.

4) 这是定理 7.2 及 3) 的一个推论. ▌

习　题

1. 设 $\{x_1,\cdots,x_m\}\subset R^n$，其中 $x_i=(\xi_{1i},\cdots,\xi_{ni})$，$i=1,\cdots,m$. 令 $\bar{x}_i=(1,\xi_{1i},\cdots,\xi_{ni})$，$i=1,\cdots,m$. 试证 $x_1,\cdots,x_m$ 仿射无关的等价条件是 R^{n+1} 中 m 个向量 $\bar{x}_1,\cdots,\bar{x}_m$ 线性无关.

2. 设 $M\subset R^n$. 试证如果 $\theta\in\mathrm{aff}M$，则 $\dim(\mathrm{aff}M)=\dim(\mathrm{span}M)$；如果 $\theta\notin\mathrm{aff}M$，则 $\dim(\mathrm{aff}M)=\dim(\mathrm{span}M)-1$.

3. 在 R^4 中，$x_1=(1,-1,2,-1)$，$x_2=(2,-1,2,0)$，$x_3=(1,0,2,0)$，$x_4=(1,0,3,1)$. 证明 x_1,x_2,x_3,x_4 仿射无关.

4. 在 R^4 中，求过 $\theta,e_1+e_2,e_1-e_2+3e_3,3e_4-e_2$ 的超平面.

5. 证明定理 2.6.

6. 证明定理 2.7.

7. 图 29 是R^2中的四个子集，试画出每一个的核。

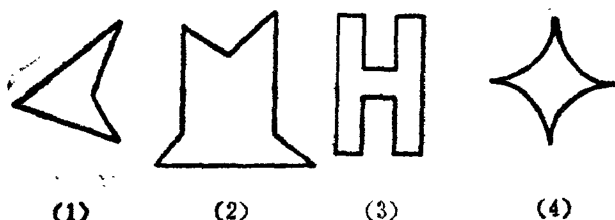


图 29

8. 设A,B是R^n中的凸集，试证$A\cap B$包含在$A\cup B$的核中。

9. 利用定义证明$C=\{x=(\xi_1,\cdots,\xi_n)\mid|\xi_1|+\cdots+|\xi_n|\leqslant 1\}$是凸集。

10. 试证具有非负元素的$n\times n$实矩阵集合是凸集。

11. 举例说明两个凸集的并不一定是凸集。

12. 试证如果x可以用两种方式表示成$x_0,x_1,\cdots,x_r$的凸组合，则$x_1-x_0,\cdots,x_r-x_0$线性相关。

13. 证明推论 3.3.1.

14. 求R^2中满足下列条件的点集的凸包：

1) $y=x^2$.

2) $y=x^2,x\geqslant 0$.

3) $y=x^3$.

4) $y=\sin x$.

15. 设A,B是R^n中的非空集合. 试证下列结论：

1) 如果$A\subset B$，则$A\subset \mathrm{co}B$，

2) 如果$A\subset B$，B是凸集，则$\mathrm{co}A\subset B$，

3) 如果$A\subset B$，则$\mathrm{co}A\subset\mathrm{co}B$，

4) $\mathrm{co}A\cup\mathrm{co}B\subset\mathrm{co}(A\cup B)$，

5) $\mathrm{co}(A\cap B)\subset\mathrm{co}A\cap\mathrm{co}B$，

6) $\mathrm{co}(\mathrm{co}A)=\mathrm{co}A$.

16．证明推论 3.8.1.

17．证明推论 3.11.1.

18．证明定理 3.13.

19．证明定理 3.15.

20．设$C\subset R^n$，试证C是凸集的充分必要条件是对于每一个 $\lambda_1\geqslant 0$，$\lambda_2\geqslant 0$，有$(\lambda_1+\lambda_2)C=\lambda_1 C+\lambda_2 C$.

21．证明定理 3.18.

22．证明定理 3.19.

23．设$C\subset R^n$，$S\subset R^n$．令

$$C_1=\bigcap_{x\in S}\bigcup_{\lambda\geq 1}\{(1-\lambda)x+\lambda C\},\quad C_2=\bigcup_{x\in S}\bigcup_{\lambda\geq 1}\{(1-\lambda)x+\lambda C\}.$$

试证：

1）如果C是凸集，则C_1是凸集.

2）如果C和S都是凸集，则C_2是凸集.

24．在R^2中，$x_1=(1,0)$，$x_2=(1,3)$，$x_3=(4,3)$，$x_4=(4,0)$，$x=\left(\frac{7}{4},\frac{5}{4}\right)$．试将$x$表示成$x_1,x_2,x_3,x_4$的凸组合．利用Caratheodory 定理证明的方式，将x表示成x_1,x_2,x_3的凸组合.

25．设A,B是R^n中的子集．试证$\mathrm{co}(A+B)=\mathrm{co}A+\mathrm{co}B$.

26．设$A\subset R^n$，φ是R^n到R^m的仿射变换．试证$\varphi(\mathrm{co}A)=\mathrm{co}\varphi(A)$.

27．设$\{x_1,\cdots,x_m\}\subset R^n$．如果$x_1,\cdots,x_m$中每一个都不是其他元素的凸组合，则称$\{x_1,\cdots,x_m\}$是凸无关的．设$m\geqslant n+2$．试证如果$\{x_1,\cdots,x_m\}$的每一个子集是凸无关的，则它也是凸无关的.

28．设R^n中的点x是$x_1,\cdots,x_m$的凸组合，每一个x_i是$y_{i1},\cdots,y_{im_i}$的凸组合，$i=1,\cdots,m$．试证x是点y_{ir_i}的凸组合，$i=1,\cdots,m$，$r_i=1,\cdots,m_i$.

29．设C是R^n中的凸集，试证$\mathrm{ri}C$是凸集.

30．设C_1,C_2是R^n中的凸集．证明$\mathrm{cl}C_1=\mathrm{cl}C_2$的充分必要条件是$\mathrm{ri}C_1=\mathrm{ri}C_2$．这个条件也等价于$\mathrm{ri}C_1\subset C_2\subset \mathrm{cl}C_1$.

31．设C是R^n中的凸集，试证每一个与$\mathrm{cl}C$相交的开集也与$\mathrm{ri}C$相交.

32．设$C\subset R^n$，试证$\mathrm{aff}C=\mathrm{aff}(\mathrm{cl}C)$.

33．举例说明定理 4.7 中C是凸集的假设是必不可少的.

34. 证明推论 4.8.2.

35. 设C_1,C_2是R^n中的凸集,试证

1) $\mathrm{cl}(\mathrm{ri}C_1+\mathrm{ri}C_2)=\mathrm{cl}(C_1+C_2)$,

2) $\mathrm{ri}(\mathrm{ri}C_1+\mathrm{ri}C_2)=\mathrm{ri}(C_1+C_2)$,

3) $\mathrm{aff}(\mathrm{ri}C_1+\mathrm{ri}C_2)=\mathrm{aff}(C_1+C_2)$.

36. 设$C\subset R^n$,试证$\mathrm{co}(\mathrm{cl}C)\subset\mathrm{cl}(\mathrm{co}C)$.

37. 设C_1,C_2是R^n中的凸集,试证

$$\mathrm{co}(\mathrm{cl}C_1\cup\mathrm{cl}C_2)\subset\mathrm{cl}(\mathrm{co}(C_1\cup C_2)).$$

38. 证明定理 5.5.

39. 证明推论 5.9.2.

40. 证明推论 5.9.3.

41. 设C是R^n中不包含原点的非空凸集. 试证存在超平面分离C和原点.

42. 设C是R^n中的闭凸集. $\bar{K}(x^*,\alpha)$,$H(x^*,\alpha)$分别是C的支撑半空间和支撑超平面. 试证当且仅当

$$\inf_{x\in C}\langle x,x^*\rangle<\max_{x\in C}\langle x,x^*\rangle$$

时,$H(x^*,\alpha)$是正常支撑超平面.

43. 设C是2维紧致凸集,试证$\mathrm{ext}C$是闭集.

44. 设C是满足下面两个条件之一的$x=(\xi_1,\xi_2,\xi_3)\in R^3$的凸包:

1) $\xi_1=\xi_2=0,\xi_3\in[-1,1]$,

2) $\xi_3=0,(\xi_1-1)^2+\xi_2^2=1$.

试说明$\mathrm{ext}C$不是闭集.

45. 设C是R^n中的闭凸集. 试证如果C的凸子集F是C的面,则$C\setminus F$是凸集. 举出反例说明上述结果的逆不成立.

46. 设C是R^n中的闭凸集,举例说明关系

$$C=\mathrm{co}(\mathrm{ext}C)$$

不成立.

47. 证明推论 6.1.1.

48. 证明推论 6.6.1.

49. 证明推论 6.10.1.

50．证明推论6.15.1.

51．设$M\subset R^n$，试证$(M^{\circ\circ})^\circ=M^\circ$.

52．设$C_i, i\in I$是R^n中包含原点的闭凸集族，I是任意指标集．试证

$$\left(\bigcap_{i\in I}C_i\right)^\circ=\text{clco}\bigcup_{i\in I}C_i^\circ.$$

53．证明推论7.3.1.

54．证明定理7.5的第二部分.

55．证明定理7.6的第二部分.

56．设$M\subset R^n$，$\lambda\neq 0$，试证$(\lambda M)^\circ=\frac{1}{\lambda}M^\circ$.

57．设$b_i\in R^n$，$i\in I$，I是任意指标集．试证

$$K=\{x\in R^n\mid\langle x,b_i\rangle\leqslant 0, i\in I\}$$

是凸锥.

58．求非负真锥的共轭锥.

59．设$\{a_i\mid i\in I\}$是R^n中的非空向量集，I是任意足标集，$K=\text{cone}\{a_i\mid i\in I\}$．求$K^*$.

60．证明推论8.1.1.

61．证明定理8.2.

62．证明定理8.3的2）.

63．证明定理8.4.

64．证明定理8.6.

65．证明定理8.15.

66．证明定理8.16.

67．证明定理8.18.

68．证明凸集C的法线锥是凸锥.

69．试证非空多面体集的每一个面都是暴露面.

70．证明定理9.3.

71．证明推论9.5.1.

72．证明定理9.13.

73．证明定理9.16.

74．证明多面体集是凸集.

75. 证明推论 9.20.2.

76. 设 C_1, C_2 是闭凸集，且至少一个有界，$C_1 \cap C_2 = \varnothing$. 试证存在向量 $x^* \in R^n$ 及 $\varepsilon > 0$，对于 $\forall x_1 \in C_1, x_2 \in C_2$ 满足

$$\langle x_1, x^* \rangle \leqslant \langle x_2, x^* \rangle - \varepsilon.$$

第二章 凸 函 数

凸函数和凸集一样，是凸分析和极值问题的重要内容．从这一章起，我们将讨论凸函数的重要性质．

以后如无特别声明，我们讨论的函数都是在广义的实轴上取值的，即可以取等于$-\infty$和$+\infty$的值．为了统一可能出现的包含$+\infty$和$-\infty$的计算，规定：

$-\infty<\alpha\leqslant+\infty$时，$\alpha+\infty=\infty+\alpha=+\infty$；

$-\infty\leqslant\alpha<+\infty$时，$\alpha-\infty=-\infty+\alpha=-\infty$；

$0<\alpha\leqslant+\infty$时，$\alpha\cdot\infty=\infty\cdot\alpha=\infty$；

$-\infty\leqslant\alpha<0$时，$\alpha\cdot\infty=\infty\cdot\alpha=-\infty$，$\alpha\cdot(-\infty)=(-\infty)\cdot\alpha=+\infty$；

$0\cdot\infty=\infty\cdot 0=0\cdot(-\infty)=(-\infty)\cdot 0=0$，$-(-\infty)=+\infty$；

$\inf\varnothing=+\infty$，$\sup\varnothing=-\infty$；

$\infty-\infty$和$-\infty+\infty$没有意义．

§ 1 凸函数的基本性质

1．定义

定义 1.1 设$f(x)$是从R^n到$[-\infty,+\infty]$的实值函数，使$f(x)$只取有限值或$-\infty$的x的集合，称为$f(x)$的有效定义域，用$\mathrm{dom}f$表示，即$\mathrm{dom}f=\{x\mid f(x)<+\infty\}$．

定义 1.2 设$f(x)$是从R^n到$[-\infty,+\infty]$的实值函数．集合$\mathrm{epi}f=\{(x,\mu)\mid x\in R^n,\mu\in R,\ \mu\geqslant f(x)\}$称为$f(x)$的上方图．

从定义可知，仅当$x\in\mathrm{dom}f$时，(x,μ)才可能是$\mathrm{epi}f$的点．因为如果$f(x)=+\infty$，则不存在这样的实数μ，满足$\mu\geqslant f(x)$．

一个函数的上方图也确定了这个函数本身．事实上，容易验

证：

$$f(x)=\inf\{\mu \mid (x,\mu)\in \mathrm{epi}f\}. \tag{1}$$

如果 R^{n+1} 中一个集合满足下面条件：若它包含点(x,μ_0)，则一定包含点(x,μ)，$\mu\geq\mu_0$，就可以认定它是一个函数的上方图，这个函数可以利用(1)式计算.

定义 1.3 设 $f(x)$是从 R^n 到$[-\infty, +\infty]$的实值函数. 如果 epi$f$ 是凸集，则称 $f(x)$是凸函数. 如果$-f(x)$是凸函数，则称 $f(x)$是凹函数. 如果 $f(x)$是凸函数也是凹函数，则称 $f(x)$是仿射函数.

注 我们没有直接定义从 R^n 的真子集 S 到$[-\infty, +\infty]$的凸函数，因为总可以将其延拓到整个 R^n 上. 如果 $f(x)$是在 $S\subset R^n$ 上定义，则

$$f_1(x)=\begin{cases} f(x) & x\in S \\ +\infty & x\notin S \end{cases}$$

在 R^n 上定义，$f(x)$和 $f_1(x)$的上方图是相同的.

图 1 是一些凸函数的上方图.

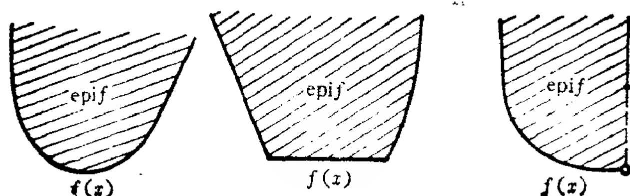


图 1

定理 1.1 设 $f(x)$是 R^n 上的凸函数，则 dom$f$ 是凸集.

证明 设 $x_1,x_2\in \mathrm{dom}f$，则存在 $\mu_1,\mu_2\in R$，满足

$$(x_1,\mu_1)\in \mathrm{epi}f,(x_2,\mu_2)\in \mathrm{epi}f.$$

因 epi$f$ 是凸集，对于 $0<\lambda<1$，有

$$(1-\lambda)(x_1,\mu_1)+\lambda(x_2,\mu_2)\in \mathrm{epi}f.$$

即 $$f[(1-\lambda)x_1+\lambda x_2]\leqslant(1-\lambda)\mu_1+\lambda\mu_2.$$

这表示$f[(1-\lambda)x_1+\lambda x_2]$不等于$+\infty$,故$(1-\lambda)x_1+\lambda x_2\in \mathrm{dom}f$, $\mathrm{dom}f$ 是凸集. ▌

定理 1.1 的逆定理不成立,即 $\mathrm{dom}f$ 是凸集，但 $f(x)$不一定是凸函数. 例如 $f(x)=x^3$，$x\in R$，它不是凸函数，但 $\mathrm{dom}f=(-\infty,+\infty)$.

称 $\mathrm{dom}f$ 的维数是 $f(x)$的维数.

定义 1.4 不等于$-\infty$且不恒等于$+\infty$的凸函数称为正常凸函数. 不是正常的凸函数称为非正常凸函数.

显然，$f(x)$是正常凸函数的等价条件是 $\mathrm{epi}f$ 非空且不包含垂直直线,或 $\mathrm{dom}f\neq\varnothing$, $x\in\mathrm{dom}f$ 时 $f(x)$有限.

例 1 函数

$$f(x)=\begin{cases}-\infty & |x|<1,\\ 0 & |x|=1,\\ +\infty & |x|>1.\end{cases}$$

是非正常凸函数.

2. 等价条件

仅仅利用定义 1.3 判断一个函数是否为凸函数是不够的，也不方便. 本节讨论一些等价条件.

定理 1.2 设 $f(x)$是 R^n 到$(-\infty, +\infty)$的实值函数. 则 $f(x)$ 是凸函数的充分必要条件是对于$\forall x,y\in R^n$, $0<\lambda<1$，有

$$f[(1-\lambda)x+\lambda y]\leqslant(1-\lambda)f(x)+\lambda f(y). \tag{2}$$

证明 设(2)成立. (x,μ), $(y,\nu)\in\mathrm{epi}f$. $(x,\mu)\neq(y,\nu)$, (z,α)是连结(x,μ)、(y,ν)的线段 l'上的一点(图 2),即

$$(z,\alpha)=(1-\lambda)(x,\mu)+\lambda(y,\nu),\quad 0<\lambda<1.$$

则

$$z=(1-\lambda)x+\lambda y, \alpha=(1-\lambda)\mu+\lambda\nu.$$

因为 $\mu\geqslant f(x)$, $\nu\geqslant f(y)$,故

$$\alpha \geqslant (1-\lambda)f(x)+\lambda f(y) \geqslant f[(1-\lambda)x+\lambda y]=f(z),$$

故$(z,\alpha)\in \mathrm{epi}f$. $\mathrm{epi}f$ 是凸集，所以 $f(x)$是凸函数.

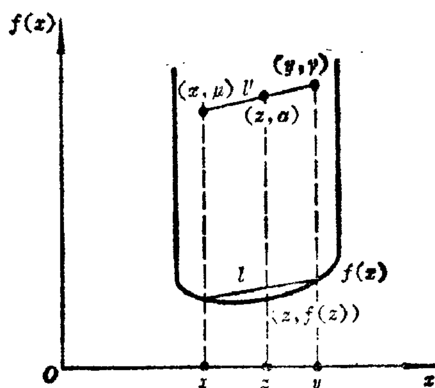


图 2

反之，设(2)不成立. 则存在 $x,y\in R^n$, $0<\lambda<1$, 使

$$f[(1-\lambda)x+\lambda y]>(1-\lambda)f(x)+\lambda f(y).$$

故点$((1-\lambda)x+\lambda y,(1-\lambda)f(x)+\lambda f(y))$属于连结点$(x,f(x))$和点$(y,f(y))$的线段 l 但不属于 $\mathrm{epi}f$, 而点$(x,f(x))$, $(y,f(y))$属于 $\mathrm{epi}f$, 所以 $\mathrm{epi}f$ 不是凸集，$f(x)$不是凸函数. ▌

定理 1.3 设 $f(x)$是 R^n 到$[-\infty,+\infty]$的实值函数. 则 $f(x)$是凸函数的充分必要条件是 $f(x)<\alpha$, $f(y)<\beta$ 时，不等式

$$f[(1-\lambda)x+\lambda y]<(1-\lambda)\alpha+\lambda\beta,\quad 0<\lambda<1 \tag{3}$$

成立.

证明 这是定理 1.2 的变形. ▌

定理 1.4 设 $f(x)$是 R^n 到$(-\infty,+\infty)$的实值函数. 则 $f(x)$是凸函数的充分必要条件是 $\forall x_i\in R^n$, $\lambda_i\geqslant 0$, $i=1,\cdots,m$, $\sum_{i=1}^{m}\lambda_i=1$, 不等式

$$f(\lambda_1x_1+\cdots+\lambda_mx_m)\leqslant\lambda_1f(x_1)+\cdots+\lambda_mf(x_m) \tag{4}$$

成立.

证明留给读者.

定义 1.5 如果(2)式中严格不等式成立，则称 $f(x)$是严格凸函数.

易知,当(2)、(3)、(4)式的不等式反向时,它们就是 $f(x)$ 为凹函数的等价条件. 对于仿射函数,(2)、(3)、(4)式变为等式.

下面讨论与微分有关的等价条件.

当$f(x)$是一元函数时容易证明下述定理.

定理 1.5 设 $f(x)$是在开区间(α,β)上的二阶连续可微的实值函数,则 $f(x)$是凸函数的充分必要条件是二阶导数 $f''(x)$在(α,β)上非负.

例 2 根据定理 1.5, 容易检验下列在 R 上定义的函数是凸函数.

1) $f(x)=e^{\alpha x}, -\infty<\alpha<+\infty$;

2) $f(x)=\begin{cases} x^p & x\geqslant 0, \\ \infty & x<0, \end{cases} 1\leqslant p<+\infty$;

3) $f(x)=\begin{cases} -x^p & x\geqslant 0, \\ +\infty & x<0, \end{cases} 0\leqslant p\leqslant 1$;

4) $f(x)=\begin{cases} x^p & x>0, \\ +\infty & x\leqslant 0, \end{cases} -\infty<p\leqslant 0$;

5) $f(x)=\begin{cases} -(\alpha^2-x^2)^{1/2}, & |x|<\alpha \\ +\infty, & |x|\geqslant\alpha \end{cases}, \alpha>0$;

6) $f(x)=\begin{cases} -\log x, & x>0 \\ +\infty, & x\leqslant 0 \end{cases}$.

设 $f(x)$是正常凸函数，$p\in R^n$, 作参数 α 的函数

$$g_{x,p}(\alpha)=f(x+\alpha p),$$

定理 1.6 当且仅当对任意的 x 和 p, $g_{x,p}(\alpha)$是 α 的凸函数时 $f(x)$是凸函数.

证明 设 $f(x)$是凸函数,则当 $0\leqslant\lambda\leqslant1$ 时

$$\begin{aligned}g_{x,p}[(1-\lambda)\alpha_1+\lambda\alpha_2]&=f\{x+[(1-\lambda)\alpha_1+\lambda\alpha_2]p\}\\&=f[(1-\lambda)(x+\alpha_1 p)+\lambda(x+\alpha_2 p)]\\&\leqslant(1-\lambda)f(x+\alpha_1 p)+\lambda f(x+\alpha_2 p)\\&=(1-\lambda)g_{x,p}(\alpha_1)+\lambda g_{x,p}(\alpha_2).\end{aligned}$$

故 $g_{x,p}(\alpha)$是凸函数.

反之,设对于任意的 $x,p,g_{x,p}(\alpha)$是 α 的凸函数. 则

$$\begin{aligned}f[(1-\lambda)x+\lambda y]&=f\{y+[(1-\lambda)\cdot1+\lambda\cdot0](x-y)\}\\&=g_{y,x-y}[(1-\lambda)\cdot1+\lambda\cdot0]\\&\leqslant(1-\lambda)g_{y,x-y}(1)+\lambda g_{y,x-y}(0)\\&=(1-\lambda)f(x)+\lambda f(y).\end{aligned}$$

故 $f(x)$是凸函数. ▌

我们用 $f'(x)$表示 $f(x)$在 x 的梯度, $f''(x)$表示 $f(x)$在 x 的二阶导数矩阵,又称 Hesse 矩阵(Hessian),分别记为

$$f'(x)=(\partial f(x)/\partial\xi_1,\cdots,\partial f(x)/\partial\xi_n),$$

$$f''(x)=\left(\frac{\partial^2 f(x)}{\partial\xi_i\partial\xi_j}\right)_{n\times n},i=1,\cdots,n,j=1,\cdots,n,$$

其中 $x=(\xi_1,\cdots,\xi_n)$.

如果 $g_{x,p}(\alpha)=f(x+\alpha p)$,利用复合函数的微分法则知

$$g'_{x,p}(\alpha)=\langle p,f'(x+\alpha p)\rangle.$$

$$g''_{x,p}(\alpha)=\langle p,f''(x+\alpha p)p\rangle.$$

定理 1.7 设 $f(x)$是 R^n 到$(-\infty,+\infty)$的可微实值函数,则以下条件彼此等价:

1) $f(x)$是凸函数,

2) 对于 $\forall x_1,x_2\in R^n$, $f(x_2)-f(x_1)\geqslant\langle x_2-x_1,f'(x_1)\rangle$,

3) 对于 $\forall x, p \in R^n, g'_{x,p}(\alpha) = \langle p, f'(x+\alpha p)\rangle$ 是 α 的不减少函数，

如果 $f(x)$ 二阶可微，还有

4) $f''(x)$ 半正定.

证明 设 1)成立. 则对 $\forall x_1 \in R^n, \forall x_2 \in R^n$，有

$f[(1-\lambda)x_1+\lambda x_2] \leqslant (1-\lambda)f(x_1)+\lambda f(x_2), 0<\lambda<1.$

于是

$$\frac{f[x_1-\lambda(x_2-x_1)]-f(x_1)}{\lambda} \leqslant f(x_2)-f(x_1).$$

令 $\lambda \downarrow 0$[注]取极限，得

$$\langle x_2-x_1, f'(x_1)\rangle \leqslant f(x_2)-f(x_1). \tag{5}$$

故 1)推出 2).

设 2)成立. 在(5)中设 $x_1 = x+\alpha_1 p, x_2 = x+\alpha_2 p, (\alpha_2 > \alpha_1)$. 得

$$g_{x,p}(\alpha_2)-g_{x,p}(\alpha_1) \geqslant (\alpha_2-\alpha_1)\langle p, f'(x+\alpha_1 p)\rangle. \tag{6}$$

而当 $x_1 = x+\alpha_2 p, x_2 = x+\alpha_1 p$ 时又有

$$g_{x,p}(\alpha_1)-g_{x,p}(\alpha_2) \geqslant (\alpha_1-\alpha_2)\langle p, f'(x+\alpha_2 p)\rangle. \tag{7}$$

由(6),(7)式得

$$\langle p, f'(x+\alpha_1 p)\rangle \leqslant \frac{g_{x,p}(\alpha_2)-g_{x,p}(\alpha_1)}{\alpha_2-\alpha_1} \leqslant \langle p, f'(x+\alpha_2 p)\rangle,$$

故 2)推出 3).

设 3)成立，对于 $\forall x, p \in R^n, g'_{x,p}(\alpha)$ 是 α 的不减少函数，则 $\alpha_2 \geqslant \alpha_1$ 时 $g'_{x,p}(\alpha_2) \geqslant g'_{x,p}(\alpha_1)$. 设 $0<\mu<1$，则

$$\begin{aligned} 0 \leqslant \mu(\alpha_2-\alpha_1)\int_0^1 \{ & g'_{x,p}[\alpha_1+t(\alpha_2-\alpha_1)] \\ & - g'_{x,p}[\alpha_1+t\mu(\alpha_2-\alpha_1)]\}dt \\ = (1-\mu) & g_{x,p}(\alpha_1)+\mu g_{x,p}(\alpha_2) \end{aligned}$$

[注] $\lambda \downarrow 0$ 意指 λ 大于零而趋于零；$\lambda \uparrow 0$ 意指 λ 小于零而趋于零.

$$-g_{x,p}[(1-\mu)\alpha_1+\mu\alpha_2],$$

即 $g_{x,p}(\alpha)$是 α 的凸函数．由定理 1.6，$f(x)$是凸函数．故 3)推出 1)．

设 $f(x)$二阶可微，$g'_{x,p}(\alpha)$是 α 的不减少函数，则

$$g''_{x,p}(\alpha)=\langle p, f''(x+\alpha p)p\rangle\geqslant 0, \tag{8}$$

故 $f''(x)$半正定，3)推出 4)．

反之，设(8)式成立，则 $g''_{x,p}(\alpha)$非负．这表明 $g'_{x,p}(\alpha)$是 α 的不减少函数，4)推出 3)．▌

例 3 设 $f(x)=\frac{1}{2}\langle x, Qx\rangle+\langle x, a\rangle+\alpha$，其中 Q 是对称的 $n\times n$ 矩阵，$a\in R^n$，$\alpha\in R$．利用定理 1.7 知 $f(x)$是凸函数的充分必要条件是 Q 半正定．

例 4 设 $x=(\xi_1,\cdots,\xi_n)$，

$$f(x)=\begin{cases}-(\xi_1\cdots\xi_n)^{1/n} & \xi_1\geqslant 0,\cdots,\xi_n\geqslant 0,\\ +\infty & \text{其他}.\end{cases}$$

证明 $f(x)$是 R^n 上的凸函数．

证明 设 $x=(\xi_1,\cdots,\xi_n)$，$z=(\eta_1,\cdots,\eta_n)$，$\xi_1>0,\cdots,\xi_n>0$．直接计算得

$$\langle z, f''(x)z\rangle=n^{-2}f(x)\left[\left(\sum_{j=1}^{n}\eta_j/\xi_j\right)^2-n\sum_{j=1}^{n}(\eta_j/\xi_j)^2\right]\geqslant 0.$$

因为 $f(x)<0$，且对任何实数 α_j，$j=1,\cdots,n$，有

$$(\alpha_1+\cdots+\alpha_n)^2\leqslant n(\alpha_1^2+\cdots+\alpha_n^2).$$

故 $f''(x)$半正定，$f(x)$是凸函数．

3．凸集和凸函数的对应

已知一个凸集 C，可以由它作出一个凸函数．这在利用 R^{n+1} 中 epi$f$ 作出 $f(x)$的公式 (1) 中已经看到．下面是这方面的另外几个例子，它们在凸分析中占有重要的地位．

1) 示性函数

$$\delta(x|C)=\begin{cases}0 & x\in C,\\ +\infty & x\notin C.\end{cases}$$

显然，C 是凸集的等价条件是 $\delta(x|C)$ 是 R^n 中的凸函数。

2) 支撑函数

$$\delta^*(x^*|C)=\sup\{\langle x^*,y\rangle|y\in C\},$$

这已在第一章§9中见过.

3) 规范函数(Minkowski 函数)

$$\nu(x|C)=\inf\{\lambda|\lambda\geqslant 0,\ x\in\lambda C\},\ C\neq\varnothing.$$

4) 距离函数

$$d(x,C)=\inf\{|x-y|\ |y\in C\}.$$

利用前面的等价条件可以验证它们是凸函数。但我们留到下一节用另外的方法检验.

已知一个凸函数 $f(x)$，有一些凸集与之对应。例如 epi f，dom f. 下面是这一类型的另一种对应.

定义 1.6 设 $f(x)$ 是凸函数，$\alpha\in[-\infty,+\infty]$，集合

$$\{x|f(x)\leqslant\alpha\}, \tag{9}$$

$$\{x|f(x)<\alpha\}, \tag{10}$$

称为水平集.

定理 1.8 水平集是凸集。

证明 只证明(9)是凸集，(10)的证明类似.

设 $x_1,x_2\in\{x|f(x)\leqslant\alpha\}$. 当 $0<\lambda<1$ 时，

$$f[(1-\lambda)x_1+\lambda x_2]\leqslant(1-\lambda)f(x_1)+\lambda f(x_2)\leqslant(1-\lambda)\alpha+\lambda\alpha=\alpha.$$

故 $(1-\lambda)x_1+\lambda x_2\in\{x|f(x)\leqslant\alpha\}$，(9)是凸集. ▌

定理 1.8 的逆定理不成立，水平集是凸集不能保证函数是凸函数，例如 $f(x)=x^3$. 水平集是凸集的函数将在最后一章专门讨论.

从几何观点看，$\{x|f(x)\leqslant\alpha\}$是 R^{n+1} 中上方图 epi f 与水平超平面$\{(x,\mu)|\mu=\alpha\}$的交在R^n上的投影．所以$\{x|f(x)\leqslant\alpha\}$可以看成 epi f 的水平截面．

推论 1.8.1 设 $f_i(x)$是 R^n 上的凸函数，$i\in I$，I 是指标集，$\alpha_i\in R$，则$\{x|f_i(x)\leqslant\alpha_i,\forall i\in I\}$是凸集．

证明留给读者．

4．正齐次凸函数

定义 1.7 设 $f(x)$是 R^n 上的实值函数．如果对于任意$x\in R^n$，$0<\lambda<+\infty$，有

$$f(\lambda x)=\lambda f(x), \tag{11}$$

则称$f(x)$是正齐次函数．

如果 $f(x)$是正齐次函数，则 $\lambda>0$ 时，有

$$\begin{aligned}\lambda\{(x,\mu)|f(x)\leqslant\mu\}&=\{(\lambda x,\lambda\mu)|f(x)\leqslant\mu\}\\&=\{(\lambda x,\lambda\mu)|\lambda f(x)\leqslant\lambda\mu\}\\&=\{(y,k)|f(y)\leqslant k\}\\&=\{(x,\mu)|f(x)\leqslant\mu\},\end{aligned}$$

所以 $f(x)$的正齐次性等价于 epi f 是 R^{n+1} 中的锥．

例 5 $f(x)=|x|$是正齐次函数．

定理 1.9 设 $f(x)$是 R^n 到$(-\infty,+\infty)$的正齐次函数，则$f(x)$是凸函数的充分必要条件是对于$\forall x,y\in R^n$，有

$$f(x+y)\leqslant f(x)+f(y). \tag{12}$$

证明 设 $f(x)$是凸函数．由(11)知

$$f(x+y)=2f\left[\frac{1}{2}(x+y)\right]\leqslant 2f\left(\frac{x}{2}\right)+2f\left(\frac{y}{2}\right)$$

$$=f(x)+f(y),$$

故(12)成立．

设(12)成立，特别地有

$$f[(1-\lambda)x+\lambda y]\leqslant f[(1-\lambda)x]+f(\lambda y)$$

$$=(1-\lambda)f(x)+\lambda f(y),$$

故 $f(x)$ 是凸函数. ▌

推论 1.9.1 设 $f(x)$ 是正齐次凸函数，则当 $\lambda_i>0$ 时，$i=1,\cdots,m$，有

$$f(\lambda_1 x_1+\cdots+\lambda_m x_m)\leqslant\lambda_1 f(x_1)+\cdots+\lambda_m f(x_m). \quad (13)$$

证明留给读者.

推论 1.9.2 设 $f(x)$ 是正齐次凸函数，则对 $\forall x\in R^n$，有

$$f(-x)\geqslant -f(x). \quad (14)$$

证明留给读者.

例 6 规范函数 $\nu(x|C)$ 是正齐次凸函数.

定义 1.8 满足(11)，(12)式的函数 $f(x)$ 称为次线性函数.

正齐次凸函数是次线性函数.

定理 1.10 正齐次正常凸函数 $f(x)$ 在子空间 L 上是线性函数的充分必要条件是对于 $\forall x\in L$，$f(-x)=-f(x)$.

实际上，只要对 L 的一个基 $b_1,\cdots,b_m$，满足 $f(-b_i)=-f(b_i)$，上面的结论就成立.

证明 设 $b_1,\cdots,b_m$ 是 L 的一个基，且 $f(-b_i)=-f(b_i)$，则 $\lambda_i\in R$ 时，$f(\lambda_i b_i)=\lambda_i f(b_i)$. 对 $x=\lambda_1 b_1+\cdots+\lambda_m b_m\in L$，有

$$\begin{aligned}
&\lambda_1 f(b_1)+\cdots+\lambda_m f(b_m)\\
&=f(\lambda_1 b_1)+\cdots+f(\lambda_m b_m)\geqslant f(x)\geqslant -f(-x)\\
&\geqslant -[f(-\lambda_1 b_1)+\cdots+f(-\lambda_m b_m)]\\
&=f(\lambda_1 b_1)+\cdots+f(\lambda_m b_m)\\
&=\lambda_1 f(b_1)+\cdots+\lambda_m f(b_m),
\end{aligned}$$

故 $f(\lambda_1 b_1+\cdots+\lambda_m b_m)=\lambda_1 f(b_1)+\cdots+\lambda_m f(b_m)$. 所以 $f(x)$ 在 L 上是线性函数.

反之，如果 $f(x)$ 在 L 上是线性函数，特别地对 $x\in L$ 有 $f(-x)=-f(x)$. ▌

§ 2 凸函数的代数运算

在实分析中的很多代数运算都是保持凸性的，本节将按对偶的方式分别讨论.

1. 复合

定理 2.1 设 $f(x)$是 R^n 到$(-\infty,+\infty)$的凸函数,$\varphi(y)$是R到$(-\infty,+\infty)$的不减少凸函数. 则复合函数 $h(x)=\varphi[f(x)]$是 R^n 上的凸函数(设 $\varphi(+\infty)=+\infty$).

证明 对于 $\forall x,y\in R^n, 0<\lambda<1$,由定理 1.2 有

$$f[(1-\lambda)x+\lambda y]\leqslant(1-\lambda)f(x)+\lambda f(y). \tag{1}$$

将(1)应用于 φ,得

$$\begin{aligned} h[(1-\lambda)x+\lambda y] &\leqslant \varphi[(1-\lambda)f(x)+\lambda f(y)] \\ &\leqslant (1-\lambda)\varphi[f(x)]+\lambda\varphi[f(y)] \\ &= (1-\lambda)h(x)+\lambda h(y). \end{aligned}$$

故 $h(x)$是凸函数. ▌

这个定理在判定具有比较复杂结构的函数的凸性时特别有用.

例 1 设 $f(x)$是正常凸函数,则$h(x)=e^{f(x)}$也是正常凸函数,因为 $\varphi(y)=e^y$ 在 R上是不减函数.

例 2 设 $f(x)$是非负凸函数,$p>1$,

$$\varphi(\xi)=\begin{cases}\xi^p & \xi\geqslant 0,\\ 0 & \xi<0.\end{cases}$$

则 $h(x)=[f(x)]^p$ 是凸函数.

例 3 设 $g(x)$是凹函数, 则 $h(x)=1/g(x)$在 $C=\{x\mid g(x)>0\}$上是凸函数,这只要对凸函数 $f=-g$ 应用于

$$\varphi(\xi)=\begin{cases}-\dfrac{1}{\xi} & \xi<0,\\ +\infty & \xi\geqslant 0.\end{cases}$$

即可得证.

2. 数乘

定义 2.1 设 $f(x)$ 是凸函数，$\lambda\geqslant 0$，称 $\lambda f(x)$ 是 $f(x)$ 的非负左乘，记为 $(\lambda f)(x)=\lambda f(x)$.

定理 2.2 若 $f(x)$ 是凸函数，则非负左乘 $\lambda f(x)$ 也是凸函数.

证明 取 $\varphi(\xi)=\lambda\xi,\xi\in R,\lambda\geqslant 0$，则 $\varphi(\xi)$ 是不减函数. 由定理2.1，$\lambda f(x)$ 是凸函数. ▌

为了定义右乘运算，先讨论下面的问题.

在§1已经看到，R^n 上的每一个凸函数 $f(x)$ 都伴随一个 R^{n+1} 中的凸集 epi f，且

$$f(x)=\inf\{\mu\mid(x,\mu)\in \text{epi } f\}. \tag{2}$$

它的反问题是，R^{n+1} 中的凸集是否一定对应 R^n 中的一个凸函数呢？如果可能，用什么方法来构造这个凸函数呢？下面的定理说明，这个函数的图形正好是 R^{n+1} 中凸集的下边界.

定理 2.3 设 F 是 R^{n+1} 中的凸集，定义

$$f(x)=\inf\{\mu\mid(x,\mu)\in F\}, \tag{3}$$

则 $f(x)$ 是 R^n 上的凸函数.

证明 设对于 $x,y\in R^n$，存在 $\alpha,\beta\in R$，满足 $f(x)<\alpha$，$f(y)<\beta$. 由 $f(x)$ 的定义，存在 $\mu,\gamma\in R,\mu<\alpha,\gamma<\beta$，使

$$(x,\mu)\in F,(y,\gamma)\in F.$$

因为 F 是凸集，故 $(1-\lambda)(x,\mu)+\lambda(y,\gamma)=((1-\lambda)x+\lambda y,(1-\lambda)\mu+\lambda\gamma)\in F$. 所以(见图 3)

$$f[(1-\lambda)x+\lambda y]\leqslant(1-\lambda)\mu+\lambda\gamma<(1-\lambda)\alpha+\lambda\beta.$$

由定理 1.3 知，$f(x)$ 是 R^n 上的凸函数. ▌

例 4 设 $C=\{(\xi_1,\xi_2)\mid\xi_1^2+\xi_2^2<1\}$. 则按(3)的方式构造的凸函数是(图 4)

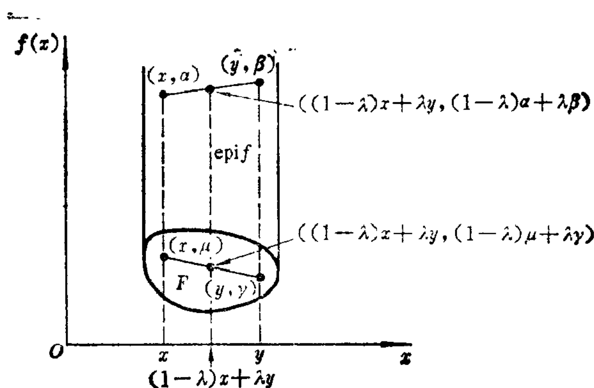


图 3

$$f(\xi_1)=\begin{cases}-(1-\xi_1^2)^{1/2} & |\xi_1|<1,\\ +\infty & |\xi_1|\geqslant 1.\end{cases}$$

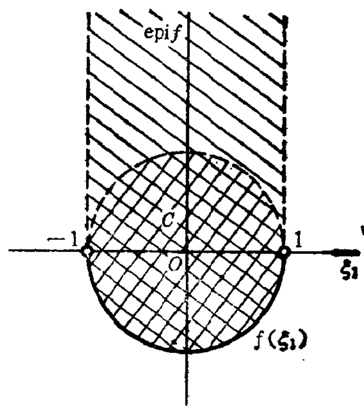


图 4

定义 2.2 设 $f(x)$ 是 R^n 上的凸函数，$0\leqslant\lambda<+\infty$，$F=\lambda\text{epi}f$. 称

$$(f\lambda)(x)=\inf\{\mu\,|\,(x,\mu)\in F\},$$

是 $f(x)$ 的非负右乘.

从定理 2.3 知，$(f\lambda)(x)$ 是凸函数. $\lambda>0$ 时，有

$$\begin{aligned}(f\lambda)(x)&=\inf\{\mu\,|\,(x,\mu)\in F\}\\&=\inf\{\lambda\gamma\,|\,(\lambda y,\lambda\gamma)\in F\}\\&=\lambda\inf\{\gamma\,|\,(y,\gamma)\in \text{epi}\,f\}\\&=\lambda\inf\left\{\gamma\,\middle|\,\left(\frac{x}{\lambda},\gamma\right)\in\text{epi}f\right\}=\lambda f(\lambda^{-1}x).\end{aligned}$$

当 $\lambda=0$，$f\not\equiv+\infty$ 时

$$F=0\cdot\text{epi}f=\{\theta\}.\quad (f0)(x)=\delta(x\,|\,\theta)$$

当 $f\equiv+\infty$ 时

$$\text{epi}\,f=\varnothing,\quad F=\varnothing.\quad 故(f0)(x)=f(x).$$

从上面的结果不难证明

定理 2.4 凸函数 $f(x)$ 是正齐次函数的充分必要条件是：$\forall\lambda>0$，$(f\lambda)(x)=f(x)$.

定义 2.3 设 $h(x)$ 是 R^n 上的凸函数，$F=\text{cone}(\text{epi } h)$. 称 $f(x)=\inf\{\mu\mid(x,\mu)\in F\}$ 是 $h(x)$ 生成的正齐次凸函数.

从定义可知，由 $h(x)$ 生成的正齐次凸函数是满足 $f(\theta)=0$，$f(x)\leqslant h(x)$ 的全体正齐次函数的最大者. 因为 $F=\bigcup\limits_{\lambda>0}\lambda\,\text{epi}h\cup\{\theta\}$，故 $h(x)\not\equiv+\infty$ 时，有

$$f(x)=\inf\{(h\lambda)(x)\mid\lambda\geqslant0\}. \tag{4}$$

当然，当 $x\neq\theta$ 或 $h(\theta)<+\infty$ 时，$\lambda=0$ 可以从下确界中略去.

例 5 R^n 中的非空凸集 C 的规范函数 $\nu(x\mid C)$ 是由 $\delta(x\mid C)+1$ 生成的正齐次凸函数. 事实上，设 $h(x)=\delta(x\mid C)+1$，则 $\lambda>0$ 时

$$\begin{aligned}(h\lambda)(x)&=\lambda h(\lambda^{-1}x)=\lambda\{\delta(\lambda^{-1}x\mid C)+1\}\\&=\lambda\delta(\lambda^{-1}x\mid C)+\lambda=\delta(x\mid\lambda C)+\lambda.\end{aligned}$$

$\lambda=0$ 时

$$(h0)(x)=\delta(x\mid\theta)=\begin{cases}0 & x=\theta,\\+\infty & x\neq\theta.\end{cases}$$

故

$$\inf\{(h\lambda)(x)\mid\lambda\geqslant0\}=\inf\{\lambda\geqslant0\mid x\in\lambda C\}=\gamma(x\mid C).$$

3. 加法和下卷积

定理 2.5 设 $f_1(x),f_2(x)$ 是 R^n 上的正常凸函数，则 $(f_1+f_2)(x)=f_1(x)+f_2(x)$ 是凸函数.

证明留给读者.

因为 $(f_1+f_2)(x)<+\infty$ 的等价条件是 $f_1(x)<+\infty$，$f_2(x)<+\infty$，故 $\text{dom}(f_1+f_2)=\text{dom } f_1\cap\text{dom} f_2$. 当 $\text{dom}(f_1+f_2)=\varnothing$ 时，表明 $(f_1+f_2)(x)$ 是非正常凸函数.

结合定理2.2，可以得到下面的推论.

推论 2.5.1 设 $f_i(x)$是 R^n 上的正常凸函数. $\lambda_i \geqslant 0$, $i=1,\cdots,m$, 则$\sum_{i=1}^{m}\lambda_i f_i(x)$是凸函数.

例 6 设 $f(x)$是R^n 上的有限凸函数, C 是非空凸集, 则

$$f(x)+\delta(x|C)=\begin{cases} f(x) & x\in C, \\ +\infty & x\notin C. \end{cases}$$

所以 $f(x)$与示性函数相加相当于限制 $f(x)$的有效定义域.

定义 2.4 设 $f_i(x)$是 R^n 上的正常凸函数, $i=1,\cdots,m$, 称

$$(f_1\square\cdots\square f_m)(x)=\inf\{f_1(x_1)+\cdots +f_m(x_m)\,|\,x_i\in R^n, x_1+\cdots+x_m=x\}, \tag{5}$$

是 $f_1(x),\cdots,f_m(x)$的下卷积.

定理 2.6 下卷积$(f_1\square\cdots\square f_m)(x)$是凸函数.

证明 设 $F_i=\operatorname{epi} f_i$, $F=F_1+\cdots+F_m$, 则 F 是 R^{n+1}中的凸集. 根据定义, 当且仅当 $(x,\mu)\in F$ 时, 存在 $x_i\in R^n$, $\mu_i\in R$, $(x_i,\mu_i)\in F_i$, 且 $x=x_1+\cdots+x_m$, $\mu=\mu_1+\cdots+\mu_m$. 于是将 F 按(3)式构造的函数正是下卷积. 由定理2.3, 下卷积是凸函数. ▌

当 $m=2$ 时, 下卷积也表示成:

$$(f\square g)(x)=\inf_y\{f(x-y)+g(y)\},$$

易知, $\operatorname{dom}(f\square g)=\operatorname{dom}f+\operatorname{dom}g$.

例 7 设 $f(x)=|x|$, $g(x)=\delta(x|C)$, C是凸集. 则函数 $(f\square g)(x)=\inf\limits_y\{|x-y|+\delta(y|C)\}=\inf\limits_{y\in C}|x-y|=d(x,C)$. 所以距离函数 $d(x,C)$是凸函数.

4. 逐点上确界和凸包

定义 2.5 设 $f_i(x)$是 R^n 上的凸函数族, $i\in I$, I 是任意指标集. 称

$$f(x)=\sup\{f_i(x)\,|\,i\in I\} \tag{6}$$

是这个函数族的逐点上确界.

定理 2.7 由(6)式定义的凸函数族的逐点上确界 $f(x)$ 是凸函数.

证明 因 $\text{epi} f_i$ 是凸集，故 $\bigcap_{i\in I}\text{epi} f_i$ 是凸集，由(6)式，有

$$\begin{aligned}\text{epi} f &= \{(x,\alpha)\mid x\in R^n, \alpha\in R, \sup f_i(x)\leqslant\alpha, i\in I\}\\ &= \{(x,\alpha)\mid x\in R^n, \alpha\in R, f_i(x)\leqslant\alpha, i\in I\}\\ &= \bigcap_{i\in I}\text{epi} f_i.\end{aligned}$$

故 $\text{epi} f$ 是凸集，于是 $f(x)$ 是凸函数. ▌

例 8 支撑函数 $\delta^*(x^*|C)=\sup\{\langle x^*,y\rangle|y\in C\}$ 是线性函数族 $\langle x^*,y\rangle$ 当 y 在 C 上变化时的逐点上确界，所以是凸函数.

例 9 设 $x=(\xi_1,\cdots,\xi_n)$. 函数 $f(x)=\max\{\xi_1,\cdots,\xi_n|x=(\xi_1,\cdots,\xi_n)\}$ 是线性函数族 $f_i(x)=\langle x,e_i\rangle$ 的逐点上确界，$i=1,\cdots,n$，其中 e_i 是 R^n 的基本单位向量. 所以 $f(x)$ 是凸函数.

定义 2.6 设 $g(x)$ 是 R^n 上的实值函数，$F=\text{co}(\text{epi} g)$. 称函数 $f(x)=\inf\{\mu|(x,\mu)\in F\}$ 是 $g(x)$ 的凸包，用 $\text{co} g(x)$ 表示.

$\text{co} g(x)$ 是不大于 $g(x)$ 的最大凸函数. 根据第一章定理 3.5，$(x,\mu)\in F$ 的等价条件是

$$\begin{aligned}(x,\mu)&=\lambda_1(x_1,\mu_1)+\cdots+\lambda_m(x_m,\mu_m)\\ &=(\lambda_1x_1+\cdots+\lambda_mx_m, \lambda_1\mu_1+\cdots+\lambda_m\mu_m),\end{aligned}$$

其中 $(x_i,\mu_i)\in\text{epi} g$，$\lambda_i\geqslant0$，$i=1,\cdots,m$，$\sum_{i=1}^m\lambda_i=1$. 故

$$\begin{aligned}f(x)&=\inf\{\mu|(x,\mu)\in F\}\\ &=\inf\Big\{\lambda_1\mu_1+\cdots+\lambda_m\mu_m|x\\ &=\lambda_1x_1+\cdots+\lambda_mx_m, g(x_i)\leqslant\mu_i, \lambda_i\geqslant0, \sum_{i=1}^m\lambda_i=1\Big\}\end{aligned}$$

$$=\inf\Big\{\lambda_1 g(x_1)+\cdots+\lambda_m g(x_m)\mid x$$

$$=\lambda_1 x_1+\cdots+\lambda_m x_m,\lambda_i\geqslant 0,\ \sum_{i=1}^{m}\lambda_i=1\ \Big\}. \tag{7}$$

其中下确界是在 x 作为 R^n的点的凸组合的全部表示式上取的，当然应避免$\infty-\infty$的发生.

定义 2.7 设$\{f_i(x)\mid i\in I\}$是 R^n上的任意函数族，I 是任意指标集. 称 $\mathrm{co}(\inf\{f_i(x)\mid i\in I\})$ 是这个函数族的凸包，用 $\mathrm{co}\{f_i(x)\mid i\in I\}$表示.

设$F=\mathrm{co}\bigcup\limits_{i\in I}\mathrm{epi}f_i$,故

$$\mathrm{co}\{f_i(x)\mid i\in I\}=\inf\{\mu\mid(x,\mu)\in F\}. \tag{8}$$

显然这是满足对于 $\forall x\in R^n$, $\forall i\in I, f_i(x)\geqslant f(x)$的凸函数 $f(x)$的最大者(不一定正常).

定理 2.8 设$\{f_i(x)\mid i\in I\}$是R^n上的正常凸函数族，$f(x)=\mathrm{co}\{f_i(x)\mid i\in I\}$,则

$$f(x)=\inf\Big\{\sum_{i\in I}\lambda_i f_i(x_i)\mid x$$

$$=\sum_{i\in I}\lambda_i x_i,\ \sum_{i\in I}\lambda_i=1,\lambda_i\geqslant 0,i\in I\Big\}, \tag{9}$$

其中下确界是在 x 作为元素 x_i 的凸组合时的全部表示式上取的，只有有限个 λ_i不为 0.

证明 根据第一章定理 3.17,$(x,\mu)\in F$ 的等价条件是

$$(x,\mu)=\sum_{i\in I}\lambda_i(x_i,\mu_i)=\Big(\sum_{i\in I}\lambda_i x_i,\ \sum_{i\in I}\lambda_i\mu_i\Big),$$

其中$(x_i,\mu_i)\in\mathrm{epi}f_i,\lambda_i\geqslant 0, i\in I,\sum\limits_{i\in I}\lambda_i=1$,且只有有限个 $\lambda_i>0$.

故由(8)得

$$f(x)=\inf\Big\{\sum_{i\in I}\lambda_i\mu_i\mid x$$

$$=\sum_{i\in I}\lambda_i x_i, f_i(x_i)\leqslant\mu_i,\ \lambda_i\geqslant 0, i\in I, \sum_{i\in I}\lambda_i=1\Big\}$$

$$=\inf\Big\{\sum_{i\in I}\lambda_i f_i(x_i)\,\Big|\, x$$

$$=\sum_{i\in I}\lambda_i x_i, \lambda_i\geqslant 0, i\in I,\ \sum_{i\in I}\lambda_i=1\Big\}.$$

即(9)成立. ▌

5. 线性变换中的象和逆象

定义 2.8 设 A 是从 R^n 到 R^m 的线性变换，$g(y)$ 是 R^m 上的凸函数，$h(x)$ 是 R^n 上的凸函数. 称 $(gA)(x)=g(Ax)$ 是在 A 中 g 的逆象. $(Ah)(y)=\inf\{h(x)\mid Ax=y\}$ 是在 A 中 h 的象.

定理 2.9 $(gA)(x)$ 是 R^n 上的凸函数，$(Ah)(y)$ 是 R^m 上的凸函数.

证明留给读者.

例 10 设 A 是投影变换，

$$A:\ x=(\xi_1,\cdots,\xi_m,\xi_{m+1},\cdots,\xi_n)\longrightarrow(\xi_1,\cdots,\xi_m),$$

$h(x)$ 是 R^n 上的凸函数，则

$$(Ah)(\xi_1,\cdots,\xi_m)=\inf_{\xi_{m+1},\cdots,\xi_n} h(\xi_1,\cdots,\xi_m,\xi_{m+1},\cdots,\xi_n).$$

由定理 2.9，$(Ah)(\xi_1,\cdots,\xi_m)$ 是凸函数.

§ 3 凸函数的闭包和连续性

线性函数是处处连续的，但一般的凸函数却并不都是如此. 研究下面的例子.

例 1 函数

$$f(x)=\begin{cases} x^2-1 & x<1, \\ 2 & x=1, \quad x\in R. \\ +\infty & x>1. \end{cases}$$

是一个凸函数(图 5)，但是它在 $\operatorname{dom} f$ 的边界点 $x=1$ 处发生间

断.

正如这一节将要看到的，凸函数的不连续性仅能发生在其有效定义域的某些边界点上. 为了消除这些不连续点，下面研究函数的闭包运算.

1. 下半连续性和闭包

定义 3.1 设 $f(x)$是在 $S\subset R^n$上定义的广义实值函数. 如果对于S中使 $x_i\xrightarrow[i\to\infty]{}x\in S$, $\lim\limits_{i\to\infty}f(x_i)$存在的每个序列$\{x_i\}$,满足

$$f(x)\leqslant\lim_{i\to\infty}f(x_i),\tag{1}$$

则称 $f(x)$在 x 下半连续. 如果$-f(x)$在 x 下半连续，称 $f(x)$在 x上半连续. 如果$f(x)$在S的每一点都是下半连续的，则称$f(x)$在S 下半连续. 如果 $f(x)$既是下半连续也是上半连续的，则称$f(x)$连续.

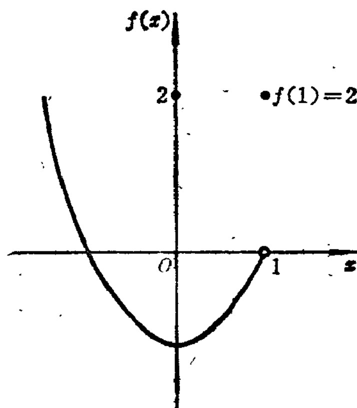


图 5

可以证明，$f(x)$在 x 下半连续的等价条件是：对于$\forall\varepsilon>0$，存在$\eta>0$，当$\forall y\in B$时，$f(x)-\varepsilon\leqslant f(y)$，其中$B=\{y\,|\,|x-y|<\eta\}$.

定理 3.1 $f(x)$在 x 下半连续的充分必要条件是

$$f(x)\leqslant\lim_{y\to x}\inf f(y).\tag{2}$$

证明 利用等价条件证明.

设 $f(x)$在 x 下半连续，则对 $\forall\varepsilon>0$，有

$$f(x)-\varepsilon\leqslant\inf_{y\in B}f(y)=\alpha(\eta)\leqslant\lim_{y\to x}\inf f(y),$$

其中 $B=\{y\,|\,|x-y|<\eta\}$. 在 $\varepsilon\to0$ 时取极限，得

$$f(x)\leqslant\liminf_{y\to x} f(y),$$

所以(2)式成立.

反之,设(2)式成立. 则对 $\forall\varepsilon>0$,存在 $\eta>0$,使

$$f(x)-\varepsilon\leqslant\inf_{y\in B} f(y),$$

其中 $B=\{y\mid |x-y|<\eta\}$. 此即表明 $f(x)-\varepsilon\leqslant f(y)$, $y\in B$. 故$f(x)$在 x 下半连续. ▌

根据定理 3.1,$f(x)$在 x 下半连续可以表示成:

$$f(x)=\liminf_{y\to x} f(y)=\lim_{\varepsilon\to 0}(\inf\{f(y)\mid |x-y|\leqslant\varepsilon\}). \qquad (3)$$

下面的定理说明下半连续在研究凸函数时的重要性.

定理 3.2 设 $f(x)$是 R^n到$[-\infty,-\infty]$的实值函数,则下列条件等价:

1) $f(x)$在 R^n下半连续,

2) $\forall\alpha\in R$,$V_\alpha=\{x\mid f(x)>\alpha\}$是开集,

3) $\forall\alpha\in R$,F_α $\{x\mid f(x)\leqslant\alpha\}$是闭集,

4) epi f 是 R^{n+1}中的闭集.

证明 设 1) 成立. $(x_i,\alpha_i)\in\mathrm{epi}f$, $(x_i,\alpha_i)\xrightarrow[i\to\infty]{}(x,\alpha)\in R^{n+1}$. 这时 $f(x_i)\leqslant\alpha_i$. 由(2),有

$$f(x)\leqslant\liminf_{i\to\infty} f(x_i)\leqslant\liminf_{i\to\infty}\alpha_i=\alpha.$$

故$(x,\alpha)\in\mathrm{epi}f$,$\mathrm{epi}f$ 是闭集,即 4)成立.

设 4)成立. $x_i\in F_\alpha$,$x_i\xrightarrow[i\to\infty]{}x$. 故 $f(x_i)\leqslant\alpha$,$(x_i,\alpha)\in\mathrm{epi}f$,且$(x_i,\alpha)\xrightarrow[i\to\infty]{}(x,\alpha)$. 但 $\mathrm{epi}f$ 是闭集,于是 $(x,\alpha)\in\mathrm{epi}f$. 即 $f(x)\leqslant\alpha$,$x\in F_\alpha$. 所以 F_α是闭集,故 3)成立.

设 3)成立. 因为 V_α是 F_α的补集,故 V_α是开集,2)成立.

设 2)成立. 如果 $x\in R^n$,$\forall\varepsilon>0$, 令 $\alpha=f(x)-\varepsilon$, 则 $x\in V_\alpha$.由 V_α是开集,存在 $\eta>0$, $V_\alpha\supset B=\{y\mid |x-y|<\eta\}$. 所以对

$\forall y\in B, f(x)-\varepsilon<f(y)$，则 $f(x)$ 在 x 下半连续，1)成立. ▌

定理 3.3 设 $\{f_i(x)\mid i\in I\}$ 是在 x 下半连续的实值函数族，I 是任意指标集，则函数族的逐点上确界 $f(x)=\sup\{f_i(x)\mid i\in I\}$ 在 x 下半连续，只要 $f(x)<+\infty$.

证明 由 $f(x)$ 的定义，对 $\forall\varepsilon>0$，存在指标 i_0，使

$$f(x)-\frac{\varepsilon}{2}\leqslant f_{i_0}(x). \tag{4}$$

而 $f_{i_0}(x)$ 在 x 下半连续，故存在 $\eta=\eta(\varepsilon,x)>0$，使当 $y\in B=\{y\mid |x-y|<\eta\}$ 时，有

$$f_{i_0}(x)-\frac{\varepsilon}{2}\leqslant f_{i_0}(y)\leqslant f(y)=\sup_I f_i(y). \tag{5}$$

结合(4)，(5)知 $y\in B$ 时 $f(x)-\varepsilon\leqslant f(y)$，即 $f(x)$ 在 x 下半连续. ▌

现在利用上方图的闭包运算定义函数的闭包运算.

定义 3.2 设 $f(x)$ 是 R^n 上的实值函数. 如果函数 $g(x)$ 满足：$\mathrm{cl}(\mathrm{epi}f)=\mathrm{epi}g$，则称 $g(x)$ 是 $f(x)$ 的下半连续包.

显然，$g(x)\leqslant f(x)$，且 $g(x)$ 是下半连续的.

定义 3.3 设 $f(x)$ 是 R^n 上不取 $-\infty$ 的凸函数. 称 $f(x)$ 的下半连续包是 $f(x)$ 的闭包，用 $\mathrm{cl}f(x)$ 表示，即

$$\mathrm{epi}(\mathrm{cl}f)=\mathrm{cl}(\mathrm{epi}f). \tag{6}$$

如果对某一个 x，$f(x)=-\infty$，则定义 $\mathrm{cl}f(x)=-\infty$. 如果满足 $\mathrm{cl}f(x)=f(x)$，则称 $f(x)$ 是闭凸函数.

定理 3.4 设 $f(x)$ 是 R^n 上的正常凸函数，则 $f(x)$ 下半连续与 $f(x)$ 是闭凸函数等价.

证明 设 $f(x)$ 是闭凸函数，即 $\mathrm{cl}f(x)=f(x)$. 则由(6)式得

$$\mathrm{epi}f=\mathrm{epi}\ \mathrm{cl}f=\mathrm{cl}\ \mathrm{epi}f,$$

即 $\mathrm{epi}f$ 是闭集. 由定理 3.2，$f(x)$ 下半连续.

反之，设 $f(x)$ 下半连续. 则由(6)式，$\mathrm{epi}f=\mathrm{cl}(\mathrm{epi}f)=\mathrm{epi}\ \mathrm{cl}f$，

故 $\mathrm{cl}f(x)=f(x)$，于是 $f(x)$ 是闭凸函数．▌

例 2　求例 1 中 $f(x)$ 的闭包．

解　$f(x)$ 是正常凸函数．易知

$$\mathrm{cl}f(x)=\begin{cases}x^2-1 & x\leqslant 1,\\ +\infty & x>1.\end{cases}$$

例 3　求函数

$$f(x)=\begin{cases}0 & \xi_1^2+\xi_2^2<R^2,\\ \lambda & \xi_1^2+\xi_2^2=R^2, \quad R、\lambda\in(0,+\infty)\\ +\infty & \xi_1^2+\xi_2^2>R^2.\end{cases}$$

的闭包．

解　$f(x)$ 是正常凸函数．易知

$$\mathrm{cl}f(x)=\begin{cases}0 & \xi_1^2+\xi_2^2\leqslant R^2,\\ +\infty & \xi_1^2+\xi_2^2>R^2.\end{cases}$$

例 2，例 3 告诉我们，只要重新定义在 $\mathrm{dom}f$ 的边界上 $f(x)$ 的某些值，就可以得到 $\mathrm{cl}f(x)$．所以闭包运算是对函数的一种合理规范化．它将某些不是下半连续的凸函数变得更加规则，从而可以利用定理 3.2 的四个性质．这在极值问题的理论和应用中都是很重要的．

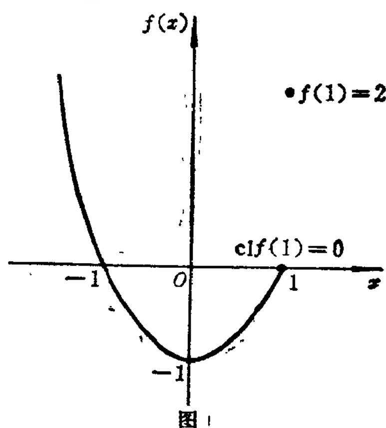


图 1

2．$\mathrm{cl}f(x)$ 和 $f(x)$ 的关系

首先讨论非正常凸函数的情形

定理 3.5　设 $f(x)$ 是非正常凸函数，则对 $\forall x\in \mathrm{ri}(\mathrm{dom}f)$，$f(x)=-\infty$．故非正常凸函数除了在有效定义域的相对边界点

外，必然是无穷大.

证明 如果 $f(x)\equiv+\infty$，结论明显.

设 $f(x)\not\equiv+\infty$. $u\in\mathrm{dom}f$，$f(u)=-\infty$，$x\in\mathrm{ri}(\mathrm{dom}f)$. 由第一章定理 4.6，存在 $\mu>1$，使 $y=(1-\mu)u+\mu x\in\mathrm{dom}f$. 故得到 $x=(1-\lambda)u+\lambda y$，$0<\lambda=\mu^{-1}<1$. 再由定理 1.3，对于任意的 $\alpha>f(u)$，$\beta>f(y)$，有

$$f(x)=f[(1-\lambda)u+\lambda y]<(1-\lambda)\alpha+\lambda\beta. \tag{7}$$

因为 $f(u)=-\infty$，$f(y)<+\infty$，在(7)式中令 $\alpha\to-\infty$，知 $f(x)=-\infty$. ▌

推论 3.5.1 下半连续的非正常凸函数没有有限值.

证明 由下半连续定义及定理 3.5 知

$$\{x\mid f(x)=-\infty\}\supset\mathrm{cl}(\mathrm{ri}\,\mathrm{dom}f).$$

但

$$\mathrm{cl}(\mathrm{ri}\,\mathrm{dom}f)=\mathrm{cl}(\mathrm{dom}f)\supset\mathrm{dom}f,$$

故 $x\in\mathrm{dom}f$ 时，$f(x)=-\infty$. 而 $x\notin\mathrm{dom}f$ 时，$f(x)=+\infty$. ▌

推论 3.5.2 设 $f(x)$ 是非正常凸函数，则 $\mathrm{cl}f(x)$ 是闭非正常凸函数. 当 $x\in\mathrm{ri}(\mathrm{dom}f)$ 时，$\mathrm{cl}f(x)=f(x)$.

证明留给读者.

现在讨论正常凸函数.

定理 3.6 设 $f(x)$ 是 R^n 上的正常凸函数，则

$$\mathrm{ri}(\mathrm{epi}f)=\{(x,\mu)\mid x\in\mathrm{ri}(\mathrm{dom}f),f(x)<\mu<+\infty\}. \tag{8}$$

证明 考虑到第一章定理 9.19 之前的说明，仅证明满维的情形，即证明

$$\mathrm{int}(\mathrm{epi}f)=\{(x,\mu)\mid x\in\mathrm{int}(\mathrm{dom}f),f(x)<\mu<+\infty\}. \tag{9}$$

因为关系“$\subset$”是明显的，只证明关系“$\supset$”. 设$(\bar{x},\bar{\mu})$满足：$\bar{x}\in\mathrm{int}(\mathrm{dom}f)$，$f(\bar{x})<\bar{\mu}$，则存在多面体 $P=\mathrm{co}\{a_1,\cdots,a_r\}\subset\mathrm{dom}f$，使 $\bar{x}\in\mathrm{int}\,P$. 令 $\alpha=\max\{f(a_i)\mid i=1,\cdots,r\}$，对$\forall x\in$

P，则 $x=\sum_{i=1}^{r}\lambda_i a_i,\lambda_i\geqslant 0,i=1,\cdots,r,\sum_{i=1}^{r}\lambda_i=1$. 因 $f(x)$ 是凸函数，故

$$f(x)\leqslant\sum_{i=1}^{r}\lambda_i f(a_i)\leqslant\sum_{i=1}^{r}\lambda_i\alpha=\alpha. \tag{10}$$

所以开集 $\{(x,\mu)\mid x\in \mathrm{int}P,\alpha<\mu<+\infty\}\subset \mathrm{epi}f$. 特别地，对于 $\forall\mu>\alpha,(\bar{x},\mu)\in \mathrm{int}(\mathrm{epi}f)$. 而 $(\bar{x},f(\bar{x}))\in \mathrm{epi}f$. 所以 $(\bar{x},\bar{\mu})\in(1-\lambda)(\bar{x},\mu)+\lambda(\bar{x},f(\bar{x})),0<\lambda<1$. 故由第一章定理 4.5，得 $(\bar{x},\bar{\mu})\in \mathrm{int}(\mathrm{epi}f)$. 即关系“$\supset$”成立. ▌

推论 3.6.1 设 $f(x)$ 是 R^n 上的正常凸函数，存在 x，满足 $f(x)<\alpha,\alpha\in R$，则存在 $\bar{x}\in \mathrm{ri}(\mathrm{dom}f)$，使 $f(\bar{x})<\alpha$.

证明 由已知条件，开半空间 $\{(x,\mu)\mid x\in R^n,\mu<\alpha\}$ 满足：

$$\{(x,\mu)\mid x\in R^n,\mu<\alpha\}\cap \mathrm{epi}f\neq\varnothing.$$

根据第一章习题 31，得

$$\{(x,\mu)\mid x\in R^n,\mu<\alpha\}\cap \mathrm{ri}(\mathrm{epi}f)\neq\varnothing,$$

即 $\{(x,\mu)\mid x\in R^n,\mu<\alpha\}\cap\{(x,\mu)\mid x\in \mathrm{ri}(\mathrm{dom}f),f(x)<\mu<+\infty\}\neq\varnothing$. 在交中取一点 $(\bar{x},\bar{\mu})$，则 $\bar{x}\in \mathrm{ri}(\mathrm{dom}f),f(\bar{x})<\bar{\mu}<\alpha$. ▌

推论 3.6.2 设 $f(x)$ 是 R^n 上的正常凸函数. 凸集 C 满足：$\mathrm{ri}C\subset \mathrm{dom}f$，且存在 $x\in \mathrm{cl}C,f(x)<\alpha,\alpha\in R$. 则存在 $\bar{x}\in \mathrm{ri}C,f(\bar{x})<\alpha$.

证明 设

$$g(x)=\begin{cases}f(x) & x\in \mathrm{cl}C,\\ +\infty & x\notin \mathrm{cl}C.\end{cases}$$

则 $\mathrm{ri}C\subset \mathrm{dom}g\subset \mathrm{cl}C$. 故 $\mathrm{ri}(\mathrm{dom}g)=\mathrm{ri}C$. 由已知条件，存在一个 $x,g(x)<\alpha$. 所以由推论 3.6.1，存在 $\bar{x}\in \mathrm{ri}(\mathrm{dom}g)$，使 $g(\bar{x})<\alpha$. 即 $\bar{x}\in \mathrm{ri}C,f(\bar{x})<\alpha$. ▌

推论 3.6.3 设 $f(x)$是 R^n 上的正常凸函数，C 是凸集，$f(x)$在 C 上有限．对任意 $x\in C, f(x)\geqslant\alpha$，则对任意 $x\in \mathrm{cl}C$，$f(x)\geqslant\alpha$．

证明留给读者．

下面是关于正常凸函数闭包的重要定理．

定理 3.7 设 $f(x)$是 R^n 上的正常凸函数．则 $\mathrm{cl}f(x)$是闭的正常凸函数．$x\notin \mathrm{cl}(\mathrm{dom}\, f)\setminus \mathrm{ri}(\mathrm{dom} f)$时，$\mathrm{cl}f(x)=f(x)$．

证明 $\mathrm{cl}f(x)$下半连续．因为 $f(x)$在 $\mathrm{dom}f$ 上有限，由推论 3.5.1，$\mathrm{cl}f(x)$是正常的．

设 $x\in \mathrm{ri}(\mathrm{dom}f)$．作垂线 $M=\{(x,\mu)\mid \mu\in R\}$，由定理 3.6，$M\cap \mathrm{ri}(\mathrm{epi}f)\neq\varnothing$．故

$$M\cap \mathrm{epi}(\mathrm{cl}f)=M\cap \mathrm{cl}(\mathrm{epi}\, f)=\mathrm{cl}(M\cap \mathrm{epi}f)=M\cap\ \mathrm{epi}f,$$

所以 $\mathrm{cl}f(x)=f(x)$．

设 $x\notin \mathrm{cl}(\mathrm{dom}f)$．因为$\mathrm{cl}(\mathrm{dom}f)\supset \mathrm{dom}(\mathrm{cl}f)\supset \mathrm{dom}f$，所以 $\mathrm{cl}f(x)=f(x)=\infty$． ▮

推论 3.7.1 设 $f(x)$是 R^n 上的正常凸函数，$\mathrm{dom}f$ 是仿射集，则 $f(x)$是闭凸函数．

证明 $\mathrm{dom}f$ 没有相对边界． ▮

从定理 3.7 可知，对于正常凸函数 $f(x)$，$\mathrm{dom}\,(\mathrm{cl}f)$ 和 $\mathrm{dom}f$ 的差别最多在 $\mathrm{dom}f$ 的边界点上，且

$$\mathrm{cl}(\mathrm{dom}(\mathrm{cl}f))=\mathrm{cl}(\mathrm{dom}f),\ \mathrm{ri}(\mathrm{dom}(\mathrm{cl}f))=\mathrm{ri}\,(\mathrm{dom}f),$$
$$\dim(\mathrm{dom}(\mathrm{cl}f))=\dim(\mathrm{dom}f).$$

3. $\mathrm{cl}f(x)$ 的计算

定理 3.8 设 $f(x)$是 R^n 上的正常凸函数，$x\in \mathrm{ri}(\mathrm{dom}f)$．则对任意 $y\in R^n$，有

$$\mathrm{cl}\ f(y)=\lim_{\lambda\uparrow 1}f[(1-\lambda)x+\lambda y]. \tag{11}$$

当 $f(x)$是非正常凸函数，且 $y\in \mathrm{cl}(\mathrm{dom}f)$时，(11) 式也成立．

证明 因为$\mathrm{cl}f(x)$下半连续，$\mathrm{cl}f(x)\leqslant f(x)$，故

$$\mathrm{cl}f(y)\leqslant\liminf_{\lambda\uparrow 1}\mathrm{cl}f[(1-\lambda)x+\lambda y]\leqslant\liminf_{\lambda\uparrow 1}f[(1-\lambda)x+\lambda y].$$

所以只需证明

$$\mathrm{cl}f(y)\geqslant\limsup_{\lambda\uparrow 1}f[(1-\lambda)x+\lambda y].\tag{12}$$

任取$\beta\geqslant\mathrm{cl}f(y)$，$\alpha>f(x)$．则

$$(y,\beta)\in\mathrm{epi}(\mathrm{cl}f)=\mathrm{cl}(\mathrm{epi}f),\quad (x,\alpha)\in\mathrm{ri}(\mathrm{epi}f).$$

故由第一章定理4.5，$0\leqslant\lambda<1$时，有

$$(1-\lambda)(x,\alpha)+\lambda(y,\beta)\in\mathrm{ri}(\mathrm{epi}f).\tag{13}$$

因为$f(x)$是凸函数，故$0\leqslant\lambda<1$时，有

$$f[(1-\lambda)x+\lambda y]<(1-\lambda)\alpha+\lambda\beta.$$

令$\lambda\to 1$取上极限，有

$$\limsup_{\lambda\uparrow 1}f[(1-\lambda)x+\lambda y]\leqslant\limsup_{\lambda\uparrow 1}[(1-\lambda)\alpha+\lambda\beta]=\beta.\tag{14}$$

由β的任意性，知(12)成立．

当$f(x)$非正常且$y\in\mathrm{cl}(\mathrm{dom}f)$时，因为$0\leqslant\lambda<1$时，$f[(1-\lambda)x+\lambda y]=-\infty$．故(11)仍成立．▌

推论 3.8.1 设$f(x)$是闭正常凸函数，则对任意$x\in\mathrm{dom}f$和$y\in R^n$，有

$$f(y)=\lim_{\lambda\uparrow 1}f[(1-\lambda)x+\lambda y].\tag{15}$$

证明 设$\varphi(\lambda)=f[(1-\lambda)x+\lambda y]$．则$\varphi(\lambda)$是$R$上的正常凸函数．$\varphi(0)=f(x)<+\infty$，$\varphi(1)=f(y)$．

因为$\{\lambda\mid\varphi(\lambda)\leqslant\alpha\}$是在连续变换：$\lambda\to(1-\lambda)x+\lambda y=z$下闭集$\{z\mid f(z)\leqslant\alpha\}$的逆象，故是闭集．根据定理3.2，$\varphi(\lambda)$下半连续．$\mathrm{dom}\varphi$是一个区间．

如果这个区间的内点在0和1之间，则取$0<\lambda'<1$，有

$$\begin{aligned}\lim_{\lambda\uparrow 1}\varphi(\lambda)&=\lim_{\lambda\uparrow 1}\mathrm{cl}\varphi(\lambda)=\lim_{\lambda\uparrow 1}\lim_{\mu\uparrow 1}\varphi[(1-\mu)\lambda'+\mu\lambda]\\&=\lim_{\mu\uparrow 1}\varphi[(1-\mu)\lambda'+\mu\cdot 1]=\mathrm{cl}\varphi(1)=\varphi(1)\end{aligned}$$

$=f(y)$.

如果这个区间的内点不在 0 和 1 之间，则 $\lim_{\lambda\uparrow 1}\varphi(\lambda)=\varphi(1)=+\infty$. ∎

例 4 设在 R^n 中，

$$f(x)=\begin{cases}-(1-|x|^2)^{1/2} & |x|\leqslant 1,\\ +\infty & |x|>1.\end{cases}$$

易知 $f(x)$ 是正常凸函数. $\mathrm{dom}f=\{x\,|\,|x|\leqslant 1\}$. 利用(11)知，在 $\mathrm{dom}\,f$ 的边界上 $\mathrm{cl}f(x)=f(x)$. 故对 $x\in R^n$, $\mathrm{cl}f(x)=f(x)$. 所以 $f(x)$ 是闭正常凸函数.

4. $\mathrm{cl}f(x)$ 与 $f(x)$ 的水平集

定理 3.9 设 $f(x)$ 是 R^n 上的正常凸函数. $\alpha\in R$, $\alpha>\inf f$. 则

1) $\mathrm{cl}\{x\,|\,f(x)\leqslant\alpha\}=\mathrm{cl}\{x\,|\,f(x)<\alpha\}=\{x\,|\,\mathrm{cl}f(x)\leqslant\alpha\}$;

2) $\mathrm{ri}\{x\,|\,f(x)\leqslant\alpha\}=\mathrm{ri}\{x\,|\,f(x)<\alpha\}=\{x\in\mathrm{ri}(\mathrm{dom}f)\,|\,f(x)<\alpha\}$;

3) 1)、2) 各集合的维数与 $\mathrm{dom}f$ 的维数相同.

证明 设 $H=\{(x,\alpha)\,|\,x\in R^n\}$ 是 R^{n+1} 中的水平超平面. 由推论 3.6.1、定理 3.6 知 $H\cap\mathrm{ri}(\mathrm{epi}f)\neq\varnothing$. 而

$$\{(x,\alpha)\,|\,f(x)\leqslant\alpha\}=H\cap\mathrm{epi}f,$$

由第一章推论 4.8.1，得

$$\begin{aligned}\mathrm{cl}\{(x,\alpha)\,|\,f(x)\leqslant\alpha\}&=H\cap\mathrm{cl}(\mathrm{epi}f)=H\cap\mathrm{epi}(\mathrm{cl}f)\\&=\{(x,\alpha)\,|\,x\in R^n\}\cap\{(x,\mu)\,|\,\mathrm{cl}f(x)\leqslant\mu\}\\&=\{(x,\alpha)\,|\,\mathrm{cl}f\{x\}\leqslant\alpha\},\end{aligned}$$

所以 $\mathrm{cl}\{x\,|\,f(x)\leqslant\alpha\}=\{x\,|\,\mathrm{cl}f(x)\leqslant\alpha\}$. (16)

$$\begin{aligned}\mathrm{ri}\{(x,\alpha)\,|\,f(x)\leqslant\alpha\}&=H\cap\mathrm{ri}(\mathrm{epi}f)\\&=\{(x,\alpha)\,|\,x\in R^n\}\cap\{(x,\mu)\,|\,x\in\mathrm{ri}(\mathrm{dom}f),f(x)<\mu\}\end{aligned}$$

$f(x)<\alpha\}$,

所 (17)

$c)\leqslant\alpha\}$.

:

$|f(x)<\alpha\}$

所以 $\mathrm{cl}\{x|$. (18)

利用第一章习题 30,得

$$\mathrm{ri}\{x|f(x)\leqslant\alpha\}=\mathrm{ri}\{x|f(x)<\alpha\}. \tag{19}$$

故 1)、2)成立. 凸集的闭包和相对内部具有相同的仿射包，故 3)成立. ▌

推论 3.9.1 设 $f(x)$是 R^n 上的闭正常凸函数，$\mathrm{dom}f$ 是相对开集. $\alpha>\inf f$, $\alpha\in R$,则

$$\mathrm{ri}\{x|f(x)\leqslant\alpha\}=\{x|f(x)<\alpha\}, \tag{20}$$

$$\mathrm{cl}\{x|f(x)<\alpha\}\quad\{x|f(x)\leqslant\alpha\}. \tag{21}$$

证明 由条件,$\mathrm{cl}f(x)=f(x)$,$\mathrm{ri}(\mathrm{dom}f)=\mathrm{dom}f$. ▌

5. 连续性

为了证明凸函数的连续性,先证明两个以后常使用的不等式.

定理 3.10 设$g(\alpha)$是 R 上的凸函数,$\alpha_0<\alpha_1<\alpha_2$,且$\alpha_0$、$\alpha_1$、$\alpha_2\in\mathrm{dom}g$,则

$$\frac{g(\alpha_2)-g(\alpha_0)}{\alpha_2-\alpha_0}\geqslant\frac{g(\alpha_1)-g(\alpha_0)}{\alpha_1-\alpha_0}, \tag{22}$$

$$\frac{g(\alpha_1)-g(\alpha_0)}{\alpha_1-\alpha_0}\leqslant\frac{g(\alpha_2)-g(\alpha_1)}{\alpha_2-\alpha_1}. \tag{23}$$

证明 设 $\lambda=\frac{\alpha_1-\alpha_0}{\alpha_2-\alpha_0}$, $1-\lambda=\frac{\alpha_2-\alpha_1}{\alpha_2-\alpha_0}$. 则 $0<\lambda<1$,且

$$\lambda\alpha_2+(1-\lambda)\alpha_0=\frac{\alpha_1-\alpha_0}{\alpha_2-\alpha_0}\alpha_2+\frac{\alpha_2-\alpha_1}{\alpha_2-\alpha_0}\alpha_0=\alpha_1.$$

因为 $g(x)$ 是凸函数，故

$$g(\alpha_1)=g[\lambda\alpha_2+(1-\lambda)\alpha_0]\leqslant\frac{\alpha_1-\alpha_0}{\alpha_2-\alpha_0}g(\alpha_2)$$

$$+\frac{\alpha_2-\alpha_1}{\alpha_2-\alpha_0}g(\alpha_0). \tag{24}$$

将(24)式两边同时减去 $g(\alpha_0)$，得

$$g(\alpha_1)-g(\alpha_0)\leqslant\frac{\alpha_1-\alpha_0}{\alpha_2-\alpha_0}[g(\alpha_2)-g(\alpha_0)],$$

故(22)式成立.

将(24)式变形为

$$\frac{\alpha_1-\alpha_0}{\alpha_2-\alpha_0}g(\alpha_1)+\frac{\alpha_2-\alpha_1}{\alpha_2-\alpha_0}g(\alpha_1)\leqslant\frac{\alpha_1-\alpha_0}{\alpha_2-\alpha_0}g(\alpha_2)$$

$$+\frac{\alpha_2-\alpha_1}{\alpha_2-\alpha_0}g(\alpha_0),$$

或

$$\frac{\alpha_2-\alpha_1}{\alpha_2-\alpha_0}[g(\alpha_1)-g(\alpha_0)]\leqslant\frac{\alpha_1-\alpha_0}{\alpha_2-\alpha_0}[g(\alpha_2)-g(\alpha_1)],$$

故(23)成立. ▌

由(22)知，差商 $\dfrac{g(\alpha)-g(\alpha_0)}{\alpha-\alpha_0}$ 是 $\alpha-\alpha_0$ 的不减少函数，这个事实以后要经常用到.

定理 3.11 设 $f(x)$ 是 R^n 上的凸函数，则 $f(x)$ 在 $\mathrm{ri}(\mathrm{dom} f)$ 的每个闭有界子集 C 上有上界.

证明 由 Heine-Borel 有限复盖定理知，C 可以被顶点在 $\mathrm{dom} f$ 内的有限个单纯形复盖. 故只要证明 $f(x)$ 在 $\mathrm{ri}(\mathrm{dom} f)$ 中的单纯形上有上界即可.

如果 $f(x)$ 是非正常的，结论自然成立.

设 $f(x)$ 是正常凸函数，下面按单纯形的维数进行归纳证明.

由定理 1.2，$f(x)$在一维单纯形(线段)上有上界：

$f[\lambda x+(1-\lambda)y]\leqslant\lambda f(x)+(1-\lambda)f(y),0<\lambda<1.$

设 $f(x)$在 $\mathrm{dom}f$ 内任一 $r-1$ 维单纯形上有上界. $S^r(x_1,\cdots,x_{r+1})$是 r 维单纯形，$x_1,\cdots,x_{r+1}\in\mathrm{ri}(\mathrm{dom}f)$.

如果 $x\in S^r(x_1,\cdots,x_{r+1}),x\neq x_{r+1}$，则

$$x=\lambda_1x_1+\cdots+\lambda_rx_r+\lambda_{r+1}x_{r+1},\lambda_i\geqslant0,i=1,\cdots,$$

$$r+1,\ \sum_{i=1}^{r+1}\lambda_i=1,$$

$\lambda_{r+1}<1$，将它改写成

$$x=\mu z+\lambda_{r+1}x_{r+1},$$

其中 $\mu=1-\lambda_{r+1},z=(\lambda_1x_1+\cdots+\lambda_rx_r)/(1-\lambda_{r+1}).x_1,\cdots,x_r$ 仿射无关，所以 z 是 $\mathrm{dom}f$ 中 $r-1$ 维单纯形的点，由归纳假设，有 $f(z)\leqslant N,N$ 与 $x_1,\cdots,x_r$ 有关. 故

$$f(x)\leqslant\mu f(z)+\lambda_{r+1}f(x_{r+1})\leqslant\max[N,f(x_{r+1})],\tag{25}$$

所以 $f(x)$在 $S^r(x_1,\cdots,x_{r+1})$上有上界. ▌

定理 3.12 设 $f(x)$是 R^n上的正常凸函数，且在 $x_0\in\mathrm{dom}f$ 的一个邻域内有上界，则 $f(x)$在 x_0 连续.

证明 不失一般性，设 $x_0=\theta,\Omega=\{x\,|\,|x|<\lambda\}$，且对于$\forall x\in\Omega,f(x)\leqslant N$. 设 $x\in\Omega$，研究函数$g(\alpha)=f(\alpha x),\alpha\in R$，在(22)中令 $\alpha_0=0,\alpha_1=\alpha>0,\alpha_2=1$，得

$$\frac{g(\alpha)-g(0)}{\alpha}\leqslant\frac{g(1)-g(0)}{1}.$$

而 $g(1)=f(x)\leqslant N,g(0)=f(\theta)\leqslant N$，所以

$$f(\alpha x)-f(\theta)\leqslant2N\alpha.\tag{26}$$

在(23)中令$\alpha_0=-1,\alpha_1=0,\alpha_2=\alpha$，得

$$\frac{g(0)-g(-1)}{0-(-1)}\leqslant\frac{g(\alpha)-g(0)}{\alpha}$$

故

$$-2N\alpha \leqslant f(\alpha x)-f(\theta). \tag{27}$$

由(26), (27)得

$$|f(\alpha x)-f(\theta)| \leqslant 2N\alpha. \tag{28}$$

任取 $\varepsilon>0$,设 $\delta=\varepsilon/(2N)<1$. $\Omega_\delta=\delta\Omega$. 设 $y\in\Omega_\delta$, 则存在 $x\in\Omega$, 满足 $y=\delta x$, 由(28), 得

$$|f(y)-f(\theta)|=|f(\delta x)-f(\theta)| \leqslant 2N\delta=\varepsilon, \tag{29}$$

所以 $f(x)$在 $x_0=\theta$ 连续. ▌

定理 3.13 设 $f(x)$是 R^n 上的正常凸函数, 则 $f(x)$在 $\mathrm{ri}(\mathrm{dom}\, f)$连续.

证明 设$\forall x_0\in \mathrm{ri}(\mathrm{dom} f)$. 则 x_0 是单纯形 $S^k(y_0,\cdots,y_k)$的内点, $y_0,\cdots,y_k\in \mathrm{dom} f$, $k=\dim(\mathrm{dom} f)$. 由定理 3.11, $f(x)$在 x_0 的一个邻域内有上界. 由定理 3.12, $f(x)$在 x_0 连续. 故 $f(x)$在 $\mathrm{ri}(\mathrm{dom}\, f)$连续. ▌

推论 3.13.1 在 R^n 上处处有限的凸函数处处连续.

证明留给读者.

例 5 设 $f(x,t)$是 $R^n\times T$ 的实值函数, 其中 T 是任意集合. 对每一个 t, $f(x,t)$是 x 的凸函数. 对每一个 x, $f(x,t)$是 t 的有界函数. 定义

$$h(x)=\sup\{f(x,t)\mid t\in T\}.$$

$h(x)$是凸函数族的逐点上确界, 因而是有限凸函数. 故 $h(x)$是 x 的连续函数.

例 6 设 $f(x)$是R^n上的处处有限凸函数, C 是 R^n 的非空凸子集, $h(x)$是 $f(x)$在平移 $C+x$ 的下确界, 则

$$\begin{aligned} h(x) &= \inf\{f(y+x)\mid y\in C\}=\inf\{f(x-z)\mid z\in -C\} \\ &= \inf_z \{f(x-z)+\delta(z\mid -C)\}=(f\Box g)(x), \end{aligned}$$

其中 $g(x)=\delta(x\mid -C)$. 所以 $h(x)$是 R^n 上的凸函数. 因为 $f(x)$处处有限, $\mathrm{dom} h=R^n$, 故 $h(x)\equiv -\infty$ 或处处有限, $h(x)$是 R^n 上的连续函数.

6. Lipschitz条件

定义 3.4 设$f(x)$是在$S\subset R^n$上的实值函数，如果存在$\alpha\geqslant 0, \alpha\in R$，满足

$$|f(y)-f(x)|\leqslant\alpha|y-x|,\quad \forall x,y\in S. \tag{30}$$

则称$f(x)$在S满足 Lipschitz 条件，简记为L-条件.

特别地，$f(x)$在S满足L-条件表示$f(x)$在S一致连续.

下面的定理是对定理 3.13 的改进.

定理 3.14 设$f(x)$是R^n上的正常凸函数，S是$\mathrm{ri}(\mathrm{dom}\,f)$中的任意闭有界子集，则$f(x)$在$S$满足$L$-条件.

证明 不失一般性，设$\dim(\mathrm{dom}\,f)=n$，则$S\subset\mathrm{int}(\mathrm{dom}\,f)$. 对任意$\varepsilon>0$，$S+\varepsilon B$是闭有界的，其中$B$是单位球. 由$B$的闭性及$S\subset\mathrm{int}(\mathrm{dom}\,f)$，故

$$\bigcap_{\varepsilon>0}(S+\varepsilon B\cap(R^n\backslash\mathrm{int}(\mathrm{dom}\,f)))=\varnothing. \tag{31}$$

所以存在一个$\varepsilon>0$，使

$$S+\varepsilon B\cap(R^n\backslash\mathrm{int}(\mathrm{dom}\,f))=\varnothing.$$

即$S+\varepsilon B\subset\mathrm{int}(\mathrm{dom}\,f)$. 由定理 3.13，$f(x)$在$S+\varepsilon B$上连续，故$f(x)$在$S+\varepsilon B$上有界. 设$N_2$、$N_1$分别是上、下界，$x,y\in S$，$x\neq y$，令$z=y+(\varepsilon/|y-x|)(y-x)$，则

$$|z|\leqslant|y|+(\varepsilon/|y-x|)|y-x|=|y|+\varepsilon, \tag{32}$$

故$z\in S+\varepsilon B$. 设$\lambda=|y-x|/(\varepsilon+|y-x|)$，$0<\lambda<1$，则

$$y=(1-\lambda)x+\lambda z.$$

因为$f(x)$是凸函数，于是

$$f(y)\leqslant(1-\lambda)f(x)+\lambda f(z)=f(x)+\lambda(f(z)-f(x)).$$

故

$$\begin{aligned}f(y)-f(x)&\leqslant\lambda(N_2-N_1)\\&=(|y-x|/(\varepsilon+|y-x|))(N_2-N_1)\\&<\alpha|y-x|,\end{aligned} \tag{33}$$

其中 $\alpha=(N_2-N_1)/\varepsilon$. 由 x,y 在 S 中的任意性,得到 $f(x)$在 S 上满足 L-条件. ▌

7. 一些函数运算的下半连续性

作为本节的最后一个小节,讨论§2 中定义的一些凸函数运算的下半连续性. 因为对于凸函数来说,下半连续性与闭性是等价的,所以在下面定理中,均使用闭性的术语.

定理 3.15 设$f_1(x),\cdots,f_m(x)$是 R^n 上的正常凸函数.

1) 如果$\mathrm{cl}\,f_i(x)=f_i(x),i=1,\cdots,m,(f_1+\cdots+f_m)(x)\not\equiv+\infty$,则$(f_1+\cdots+f_m)(x)$是闭正常凸函数.

2) 如果$f_i(x)$不全是闭凸函数,但$\bigcap\limits_{i=1}^{m}\mathrm{ri}(\mathrm{dom}\,f_i)\neq\varnothing$,则

$$\mathrm{cl}(f_1+\cdots+f_m)(x)=\mathrm{cl}\,f_1(x)+\cdots+\mathrm{cl}\,f_m(x). \tag{34}$$

证明 1) 在§2 中已经知道,$(f_1+\cdots+f_m)(x)$是正常凸函数,故只证明它是闭凸函数.

设 $f(x)=(f_1+\cdots+f_m)(x)$, $x\in\mathrm{ri}(\mathrm{dom}f)=\mathrm{ri}\left(\bigcap\limits_{i=1}^{m}\mathrm{dom}\,f_i\right)$.

对于$\forall\, y\in R^n$,由定理 3.8,有

$$\begin{aligned}\mathrm{cl}\,f(y)&=\lim_{\lambda\uparrow1}f[(1-\lambda)x+\lambda y]\\&=\sum_{i=1}^{m}\lim_{\lambda\uparrow1}f_i[(1-\lambda)x+\lambda y]\\&=\sum_{i=1}^{m}\mathrm{cl}\,f_i(y)=(f_1+\cdots+f_m)(y)\\&=f(y),\end{aligned} \tag{35}$$

故$(f_1+\cdots+f_m)(x)$是闭凸函数.

2) 由第一章定理 4.8,知

$$\bigcap_{i=1}^{m} \mathrm{ri}(\mathrm{dom} f_i)=\mathrm{ri}(\mathrm{dom}\ f).$$

因为$\bigcap_{i=1}^{m} \mathrm{ri}(\mathrm{dom} f_i)\neq\varnothing$,故可取 $x\in\mathrm{ri}(\mathrm{dom}\ f)$，再利用定理 3.8 就可类似地推出(34)是成立的. ▌

定理 3.16 设 $f_i(x), i\in I$ 是 R^n 上的正常凸函数族, I 是任意指标集, $f(x)=\sup\{f_i(x)\mid i\in I\}$.

1) 如果 $f(x)$在某些点有限, $\mathrm{cl}\ f_i(x)=f_i(x), i\in I$, 则 $f(x)$ 是闭正常凸函数.

2) 如果 $f_i(x)$不全是闭凸函数，但存在$\bar{x}\in\bigcap_{i\in I} \mathrm{ri}(\mathrm{dom}\ f_i)$, $f(\bar{x})<+\infty$,则$\mathrm{cl} f(x)=\sup\{\mathrm{cl} f_i(x)\mid i\in I\}$.

证明 1) 由§2 知 $f(x)$是正常凸函数. 因为 $\mathrm{epi} f_i$ 是闭集, $\mathrm{epi} f=\bigcap_{i\in I}\mathrm{epi} f_i$, 故 $\mathrm{epi}\ f$ 是闭集, 所以 $f(x)$是闭正常凸函数.

2) 由第一章定理 4.8, 有

$$\mathrm{epi}(\mathrm{cl} f)=\mathrm{cl}\ \ \mathrm{epi}\, f=\mathrm{cl}\bigcap_{i\in I}\mathrm{epi}\, f_i=\bigcap_{i\in I}\mathrm{cl}\ \ \mathrm{epi} f_i$$

$$=\bigcap_{i\in I}\mathrm{epi}(\mathrm{cl} f_i)=\mathrm{epi}\sup\{\mathrm{cl} f_i\mid i\in I\}, \qquad (36)$$

所以

$$\mathrm{cl} f(x)=\sup\{\mathrm{cl} f_i(x)\mid i\in I\}. ▌$$

定理 3.17 设 A 是从 R^n 到 R^m 的线性变换, $g(y)$是 R^m 上使$(gA)(x)\not\equiv+\infty$的正常凸函数.

1) 如果$\mathrm{cl} g(y)=g(y)$, 则$\mathrm{cl}(gA)(x)=(gA)(x)$.

2) 如果g 不是闭凸函数，但存在 x, 使 $Ax\in\mathrm{ri}(\mathrm{dom} g)$，则 $\mathrm{cl}(gA)(x)=((\mathrm{cl} g)A)(x)$.

证明 1) 由定理 2.9，$(gA)(x)$ 是正常凸函数．$\mathrm{epi}(gA)$是在线性变换B：$(x,\mu)\longrightarrow(Ax,\mu)$下$\mathrm{epi}\,g$的逆象，故当 $\mathrm{cl}\,g(y)=g(y)$时，$\mathrm{cl}(gA)(x)=(gA)(x)$．

2) 由第一章定理 4.10 及本章定理 3.6，得

$$
\begin{aligned}
\mathrm{epi}(\mathrm{cl}(gA)) &= \mathrm{cl}\ \mathrm{epi}(gA)=\mathrm{cl}(B^{-1}\mathrm{epi}\ g)\\
&= B^{-1}(\mathrm{cl}\ \mathrm{epi}g)=B^{-1}\mathrm{epi}(\mathrm{cl}\ g)\\
&= \mathrm{epi}(\mathrm{cl}g)A,
\end{aligned}
$$

所以

$$
\mathrm{cl}(gA)(x)=((\mathrm{cl}g)A)(x). \blacksquare
$$

§ 4 共轭函数

这一节及§5主要讨论凸函数的对偶理论．这是凸分析的一个重要的研究领域，它在极值问题中的作用正越来越受到重视．

1．共轭函数的概念

设$f(x)$是R^n上的闭凸函数，故 $\mathrm{epi}f$ 是 R^{n+1} 中的闭凸集．计算$\mathrm{epi}f$的支撑函数：

$$
\delta^*((x^*,\mu^*)|\mathrm{epi}f)=\sup\{\langle x,x^*\rangle+\mu\mu^*|(x,\mu)\in\mathrm{epi}f\},
$$

如果 $\mu^*>0$，对于固定的 x，μ 可以是大于 $f(x)$ 的任意实数．故 $\delta^*((x^*,\mu^*)|\mathrm{epi}f)=+\infty$．在$\mu^*\leqslant 0$时，后面将会看到，只要计算在 $\mu^*=-1$ 时的值就可以了．设 $\mu^*=-1$，则有

$$
\begin{aligned}
\delta^*((x^*,-1)|\mathrm{epi}f) &= \sup\{\langle x,x^*\rangle-\mu|\mu\geqslant f(x)\}\\
&= \sup\{\langle x,x^*\rangle-f(x)|x\in R^n\}. \qquad (1)
\end{aligned}
$$

由(1)式，便得出下述共轭函数的概念．

定义 4.1 设$f(x)$是R^n上的凸函数，则称

$$
f^*(x^*)=\sup\{\langle x,x^*\rangle-f(x)|x\in R^n\} \qquad (2)
$$

为 $f(x)$的共轭函数．

定理 4.1 设 $f(x)$是 R^n 上的凸函数，则 $f^*(x^*)$ 是闭凸函数．

证明 对于固定的 x, $h_x(x^*)=\langle x,x^*\rangle-f(x)$是仿射函数，因而是闭凸函数．而由(2)式，$f^*(x^*)$ 是函数族 $h_x(x^*)$关于 x 的逐点上确界，$x\in R^n$，故

$$\operatorname{epi} f^*=\bigcap_{x\in R^n}\operatorname{epi} h_x.$$

因为 epi h_x 是闭凸集，故 epi f^* 是闭凸集，所以 $f^*(x^*)$是闭凸函数．▌

定理 4.2(Fenchel不等式) 设 $f(x)$是R^n 上的凸函数，则不等式

$$f(x)+f^*(x^*)\geqslant\langle x,x^*\rangle,\forall x,x^*\in R^n \tag{3}$$

成立．

证明 由(2)，得

$$f^*(x^*)\geqslant\langle x,x^*\rangle-f(x),\forall x,x^*\in R^n, \tag{4}$$

所以(3)成立．▌

实际上，$f^*(x^*)$，$f(x)$是满足不等式类

$$f(x)+g(y)\geqslant\langle x,y\rangle,\forall x,y\in R^n. \tag{5}$$

的“最优解”．“最优解”是指不等式(5)不能再变紧的函数对$(\bar{f},\bar{g})$，即如果令W表示使不等式(5)成立的函数对(f,g)的集合，对于任意$(f',g')\in W$，$f'\leqslant\bar{f}$，$g'\leqslant\bar{g}$，则一定有

$$f'=\bar{f},\quad g'=\bar{g}.$$

显然，$(f,g)\in W$的充分必要条件是

$$g(y)\geqslant\sup_x\{\langle x,y\rangle-f(x)\}=f^*(y),\forall y\in R^n. \tag{6}$$

或等价地，有

$$f(x)\geqslant\sup_y\{\langle x,y\rangle-g(y)\}=g^*(x),\forall x\in R^n. \tag{7}$$

因此，由(6)，(7)式可知，W中的最优函数对恰好是使 $g=f^*$，$f=g^*$的(f,g)，所以最优的不等式对应互为共轭的闭凸函数对．

图 7 是共轭函数的几何解释．设

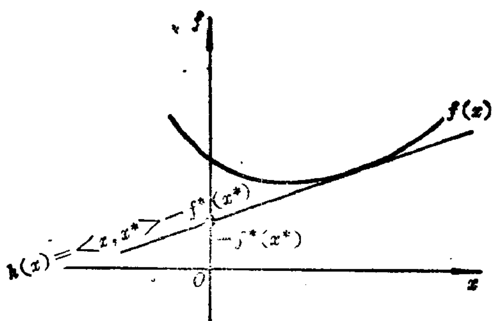


图 7

$$h(x)=\langle x,x^*\rangle-f^*(x^*),$$

因为$f(x)\geqslant\langle x,x^*\rangle-f^*(x^*)$，故$h(x)$是满足$f(x)\geqslant h(x)$的仿射函数，且在某一点，有$h(x)=f(x)$，即$h(x)$是系数为$x^*$，且满足$f(x)\geqslant h(x)$的最大仿射函数．它与$f$轴的交点是$(\theta,-f^*(x^*))$，$x^*$变化则$-f^*(x^*)$的值也上下变化，$h$与$f$相切的点（如果存在）则左右移动．

共轭函数也有合理的经济解释．工厂生产数量分别为$x_1,\cdots,x_n$的n种产品，成本由$f(x)$表示，$x=(x_1,\cdots,x_n)$．这些产品的单位价格分别为$\xi_1,\cdots,\xi_n$，盈利是$\sum_{i=1}^{n}x_i\xi_i-f(x)$．工厂的决策就是选择$x_1,\cdots,x_n$的值，使盈利最大．最大盈利作为价格的函数就是$f^*(x^*)$，$x^*=(\xi_1,\cdots,\xi_n)$．

2．基本性质

定理 4.3　设$f(x)$是R^n上的闭正常凸函数，则

$$f(x)=f^{**}(x). \tag{8}$$

其中$f^{**}(x)=(f^*)^*(x)=\sup\{\langle x,x^*\rangle-f^*(x^*)\mid x^*\in R^n\}$．

证明　由 Fenchel 不等式(3)，有

$$f(x)\geqslant\langle x,x^*\rangle-f^*(x^*),$$

则

$$f(x)\geqslant\sup\{\langle x,x^*\rangle-f^*(x^*)\mid x^*\in R^n\}=f^{**}(x),\tag{9}$$

由此还可得出 $\mathrm{dom}\,f\subset\mathrm{dom}\,f^{**}$.

为了证明反向的不等式，分两种情况讨论.

1) 设 $x_0\in\mathrm{dom}f$，任取 $\beta<f(x_0),\varepsilon>0$,令

$$P_\varepsilon=\{(y,\nu)\mid y\in x_0+\varepsilon B,\nu\leqslant\beta\},\tag{10}$$

其中B是单位球,因为$\mathrm{epi}f$是闭集,故当ε充分小时,$P_\varepsilon\cap\mathrm{epi}f=\varnothing$(其细节留给读者),见图8.

P_ε和 $\mathrm{epi}f$ 可以分离,由第一章定理 5.5,在 R^{n+1} 中存在非零向量(x^*,μ^*),对$\forall(x,\mu)\in\mathrm{epi}f,(y,\nu)\in P_\varepsilon$,满足

$$\langle y,x^*\rangle+\nu\mu^*\geqslant\langle x,x^*\rangle+\mu\mu^*.\tag{11}$$

因为ν可以取小于β的任意实数,所以由(11)式可知,$\mu^*\leqslant0$.

下面证明$\mu^*<0$. 假设$\mu^*=0$,则由(11)式知,对 $\forall x\in\mathrm{dom}f$, $\forall y\in x_0+\varepsilon B$,有

$$\langle y,x^*\rangle\geqslant\langle x,x^*\rangle.\tag{12}$$

令 $x=x_0$，得

$$\langle y-x_0,x^*\rangle\geqslant0.$$

由于y在 $x_0+\varepsilon B$ 中任意取值,故$x^*=\theta$,这与(x^*,μ^*)是非零向量矛盾,所以 $\mu^*<0$.

故可设 $\mu^*=-1$. 这时(11)式变成

$$\langle y,x^*\rangle-\nu\geqslant\langle x,x^*\rangle-\mu,\tag{13}$$

其中$\nu\leqslant\beta,y\in x_0+\varepsilon B,\mu\geqslant f(x)$.

令$\nu=\beta,\mu=f(x),y=x_0$,则由(13)式得

$$\langle x_0,x^*\rangle-\beta\geqslant\sup_x\{\langle x,x^*\rangle-f(x)\}=f^*(x^*),$$

或

$$\beta\leqslant\langle x_0,x^*\rangle-f^*(x^*).$$

所以，有

$$\beta\leqslant\sup_{x^*}\{\langle x_0,x^*\rangle-f^*(x^*)\}=f^{**}(x_0).\tag{14}$$

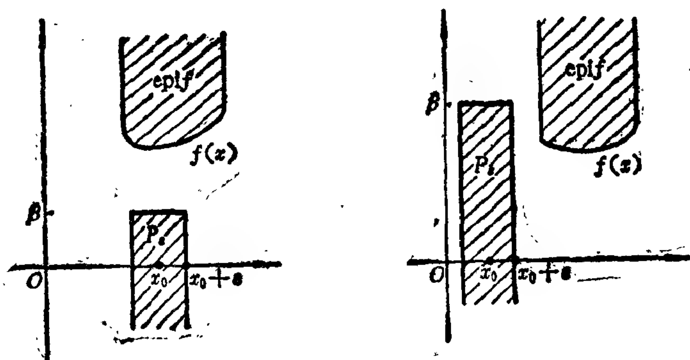


图 8

由β的任意性，得到

$$f(x_0)\leqslant f^{**}(x_0).$$

故当 $x\in \mathrm{dom}\, f$ 时，$f(x)=f^{**}(x)$.

2) 设 $x_0\notin \mathrm{dom}\, f$. 这时 $f(x_0)=+\infty$. 用与 1) 类似的讨论可知，对于任意的 β，当 ε 充分小时，$\mathrm{epi}\, f$ 与 P_ε 不相交，故(11)式成立. 这时又可分为两种情形讨论.

如果对于充分大的 β(注意到$f(x_0)=+\infty,\beta<f(x_0)$)，$\mu^*<0$. 这时(14)式成立，它表示 $f^{**}(x_0)=+\infty$.

如果对于某一个 β，$\mu^*=0$. 这时 $x^*\neq\theta$，则由(11)式，对于 $\forall x\in \mathrm{dom}\, f,\forall y\in x_0+\varepsilon B$，有

$$\langle y,x^*\rangle\geqslant\langle x,x^*\rangle. \tag{15}$$

将(15)式两边分别减去$\langle x_0,x^*\rangle$，又得到

$$\langle y-x_0,x^*\rangle\geqslant\langle x,x^*\rangle-\langle x_0,x^*\rangle \tag{16}$$

将(16)式左边在 $x_0+\varepsilon B$ 上取极小值，右边在 $\mathrm{dom}\, f$ 上取上确界，得

$$-\varepsilon|x^*|\geqslant\sup\{\langle x,x^*\rangle-\langle x_0,x^*\rangle\,|\,x\in \mathrm{dom} f\}. \tag{17}$$

因为$f(x)$是正常凸函数，故存在一点 x_1，$f(x_1)$是有限值. 对 x_1，令$\beta<f(x_1)$. 应用与 1)类似的推证，在(13)式中令$\nu=\beta,y=$

$x_1, \mu = f(x)$，得

$$-\beta + \langle x_1, x^* \rangle \geqslant -f(x) + \langle x, x^* \rangle,$$

即$-\beta + \langle x_1, x^* \rangle \geqslant f^*(x^*)$，这表示存在$x_1^*$，$f^*(x_1^*)$有限．故由定义 4.1，在 $\alpha > 0$ 时，有

$$\begin{aligned} f^*(x_1^* + \alpha x^*) &= \sup\{\langle x, x_1^* + \alpha x^* \rangle - f(x) \mid x \in R^n\} \\ &\leqslant f^*(x_1^*) + \alpha \sup_{x \in \mathrm{dom} f} \langle x, x^* \rangle. \end{aligned} \tag{18}$$

由(17)，(18)及 Fenchel 不等式，得

$$\begin{aligned} f^{**}(x_0) &\geqslant \langle x_0, x_1^* + \alpha x^* \rangle - f^*(x_1^* + \alpha x^*) \\ &\geqslant \langle x_0, x_1^* \rangle - f(x_1^*) + \alpha[\langle x_0, x^* \rangle - \sup_{x \in \mathrm{dom} f} \langle x, x^* \rangle] \\ &\geqslant \langle x_0, x_1^* \rangle - f(x_1^*) + \alpha \varepsilon |x^*|. \end{aligned}$$

当$\alpha \to +\infty$时，上式右边趋于$+\infty$，故$f^{**}(x_0) = +\infty$。

所以，$x \notin \mathrm{dom}\, f$ 时，$f(x) = f^{**}(x)$．▌

在定理 4.3 的证明中，当$x_0 \in \mathrm{dom} f$ 的情形下，只要$\varepsilon > 0$ 充分小，就可保证 P_ε 与 epi f 不相交，从而得到所要证明的结论．所以，只要 $f(x)$在 x_0 处下半连续，则定理的结论在 x_0处是成立的．故有下面的定理：

定理 4.4 设 $f(x)$是 R^n中的正常凸函数，$x_0 \in \mathrm{dom}\, f$，$f(x)$在 x_0 下半连续，则

$$f(x_0) = f^{**}(x_0).$$

定理 4.5 设 $f(x)$是 R^n上的闭正常凸函数，则 $f^*(x^*)$是闭正常凸函数．

证明 根据定理 4.1，只需证明$f^*(x^*)$是正常的即可．

因为 $f(x)$是正常的，故存在一点 $x_1 \in \mathrm{dom}\, f$，$f(x_1)$有限．由(3)式，对 $\forall x^* \in R^n$，有

$$f^*(x^*) \geqslant \langle x_1, x^* \rangle - f(x_1),$$

所以 $f^*(x^*)$ 不取$-\infty$．在定理 4.3 的 2) 中已经证明，存在一个x_1^*，$f^*(x_1^*)$的值有限，于是 $f^*(x^*) \neq +\infty$，故 $f^*(x^*)$是正常的．▌

由于$(\mathrm{cl}\,f)^*(x^*)=f^*(x^*)$，所以当$f(x)$是一般的凸函数(不一定闭)时，下面的结果成立：

$$f^{**}=(\mathrm{cl}\,f)^{**}=\mathrm{cl}\,f. \tag{19}$$

对于一般的凸函数，下面的公式也成立：

$$f^*(x^*)=\sup\{\langle x,x^*\rangle-f(x)\mid x\in \mathrm{ri}(\mathrm{dom}f)\}. \tag{20}$$

定理 4.6 设$f(x)$是R^n上的凸函数．则下面的结果成立：

1) 如果$\Phi(x)=f(x)+\alpha$，$\alpha\in R$，则$\Phi^*(x^*)=f^*(x^*)-\alpha$.

2) 如果$\Phi(x)=f(x+z)$，$z\in R^n$，则$\Phi^*(x^*)=f^*(x^*)-\langle x^*,z\rangle$.

3) 如果$\Phi(x)=f(\alpha x)$，$\alpha\neq 0$，则$\Phi^*(x^*)=f^*\left(\frac{x^*}{\alpha}\right)$.

4) 如果$\Phi(x)=\alpha f(x)$，$\alpha>0$，则$\Phi^*(x^*)=\alpha f^*\left(\frac{x^*}{\alpha}\right)$.

证明留给读者.

如果$f_1(x)$，$f_2(x)$是R^n上的凸函数，$f_1(x)\geqslant f_2(x)$，由定义4.1可以得到

$$f_1^*(x^*)\leqslant f_2^*(x^*). \tag{21}$$

定理 4.7 设$f(x)$是R^n中的闭正齐次凸函数，则

$$f^*(x^*)=\delta(x^*\mid \mathrm{dom}f^*). \tag{22}$$

证明 先证明$f(\theta)=0$.

由定理1.9知，对于$\forall x,y\in R^n$，有

$$f(x+y)\leqslant f(x)+f(y).$$

特别地，令$x\in \mathrm{ri}(\mathrm{dom}\,f)$，$y=\theta$，则有

$$f(\theta)+f(x)\geqslant f(x),$$

故$f(\theta)\geqslant 0$.

因为$f(x)$是闭凸函数，且由推论3.8.1，有

$$f(\theta)=\mathrm{cl}f(\theta)=\lim_{\lambda\uparrow 1}f[(1-\lambda)x]=\lim_{\lambda\uparrow 1}(1-\lambda)f(x)=0.$$

故$f(\theta)=0$.

计算 $f(x)$的共轭函数：

$$f^*(x^*)=\sup\{\langle x,x^*\rangle-f(x)\mid x\in R^n\}\geqslant -f(\theta)=0. \quad (23)$$

如果存在$x_1^*\in R^n$，$f^*(x_1^*)>0$，则存在 $x_1\in R^n$，使

$$\langle x_1,x_1^*\rangle-f(x_1)>0.$$

从而对$\alpha>0$时，有

$$f^*(x_1^*)\geqslant\sup_{\lambda>\theta}\{\langle\lambda x_1,x_1^*\rangle-f(\lambda_1 x_1)\}$$

$$=\sup_{\lambda>0}\lambda\{\langle x_1,x_1^*\rangle-f(x_1)\}=+\infty.$$

所以$f^*(x^*)$只取两个值：0 和$+\infty$，故

$$f^*(x^*)=\delta(x^*\mid\mathrm{dom}f^*)=\begin{cases}0 & x^*\in\mathrm{dom}f^*,\\ +\infty & x^*\notin\mathrm{dom}f^*.\end{cases}$$

例 1 设$f(x)=e^x$，$x\in R$，求$f^*(x^*)$.

解 由定义 4.1，有

$$f^*(x^*)=\sup_x\{xx^*-e^x\}. \quad (24)$$

如果$x^*<0$，取 $x<0$. 当 x 的绝对值充分大时，可使 xx^*-e^x 充分大，故(24)式中的上确界是$+\infty$. $x^*=0$ 时，(24)式中的上确界显然是 0. 如果 $x^*>0$，利用初等微积分知识计算知(24)式中的上确界是 $x^*\log x^*-x^*$. 故有下面的结果：

$$f^*(x^*)=\begin{cases}x^*\log x^*-x^* & x^*>0,\\ 0 & x^*=0,\\ +\infty & x^*<0.\end{cases}$$

当然，$f^*(0)$也可作为$x^*\downarrow 0$ 时$f^*(x^*)$的极限而得到. 不难验证，$f^{**}=f$.

例 2 设$\Phi(x)=\langle a,x\rangle-b$，其中 x，$a\in R^n$，$b\in R$. 求 $\Phi^*(x^*)$

解 设 $f(x)=\langle a,x\rangle$，则

$$f^*(x^*)=\sup_x\{\langle x,x^*\rangle-\langle a,-x\rangle\}=\begin{cases}0 & x^*=a,\\ +\infty & x^*\neq a.\end{cases}$$

根据定理 4.6 的 1)，得

$$\Phi^*(x^*)=\begin{cases} b & x^*=a, \\ +\infty & x^*\neq a. \end{cases}$$

例 3 设$\Phi(x)=\frac{1}{2}\langle x,x\rangle, x\in R^n$. 求$\Phi^*(x^*)$.

解 设$f(x)=\langle x,x\rangle$，则

$$f^*(x^*)=\sup_x\{\langle x,x^*\rangle-\langle x,x\rangle\}$$

$$=\frac{1}{4}\langle x^*,x^*\rangle.$$

根据定理 4.6 的 4)，得

$$\Phi^*(x^*)=\frac{1}{2}\cdot\frac{1}{4}\langle 2x^*,2x^*\rangle=\frac{1}{2}\langle x^*,x^*\rangle.$$

例 4 证明当且仅当 $f(x)=\frac{1}{2}\langle x,x\rangle$ 时，恒等式 $f^*=f$ 成立.

证明 由例 3，当$f(x)=\frac{1}{2}\langle x,x\rangle$时，$f^*=f$.

反之，设 f 满足恒等式 $f^*=f$，且 $w=\frac{1}{2}\langle x, x\rangle$. 由 Fenchel 不等式，有

$$\langle x,x\rangle\leqslant f(x)+f^*(x)=2f(x), \tag{25}$$

由(25)得$f(x)\geqslant w(x)$，于是$f^*\leqslant w^*$. 因为$f=f^*, w=w^*$，故$f=f^*\leqslant w^*=w$，所以 $f(x)=w(x)=\frac{1}{2}\langle x,x\rangle$.

例 5 设

$$f(x)=\begin{cases} -\log x & x>0, \\ +\infty & x\leqslant 0. \end{cases} \quad x\in R.$$

求 $f^*(x^*)$.

解 由定义，有

$$f^*(x^*)=\sup_{x>0}\{xx^*+\log x\}.$$

$x^*\geqslant 0$ 时，上确界为 $+\infty$. $x^*<0$ 时，经计算知，上确界为 $-1-\log(-x^*)$，故

$$f^*(x^*)=\begin{cases}-1-\log(-x^*) & x^*<0,\\ +\infty & x^*\geqslant 0.\end{cases}$$

例 6 设

$$f(x)=\begin{cases}-\dfrac{1}{2}-\log x & x>0,\\ +\infty & x\leqslant 0.\end{cases}$$

求 $f^*(x^*)$.

解 由定理 4.6 的 1) 及例 5 的结果知

$$f^*(x^*)=\begin{cases}-\dfrac{1}{2}-\log(-x^*) & x^*<0,\\ +\infty & x^*\geqslant 0.\end{cases}$$

显然，对于本例中的 $f(x)$，关系式 $f^*(x)=f(-x)$ 成立.

例 7 设 L 是 R^n 中的子空间，$f(x)=\delta(x\,|\,L)$，求 $f^*(x^*)$.

解 由定义 4.1 知

$$f^*(x^*)=\sup_{x}\{\langle x,x^*\rangle-\delta(x\,|\,L)\}=\sup_{x\in L}\{\langle x,x^*\rangle\}$$

$$=\begin{cases}0 & x^*\in L^{\perp}\\ +\infty & x^*\notin L^{\perp}\end{cases}=\delta(x^*\,|\,L^{\perp}).$$

因为 $f^{**}=f$，而 $L^{\perp\perp}=L$，在这个意义上，子空间的正交对应可以看成凸函数共轭对应的特殊情形.

例 8 设 $h(x)$ 是 R^n 中的凸函数，而

$$f(x)=h[A(x-a)]+\langle x,a^*\rangle+\alpha,$$

其中 A 是从 R^n 到 R^n 的一对一线性交换，$a,a^*\in R^n$，$\alpha\in R$. 求 $f^*(x^*)$.

解 设 $y=A(x-a)$，则 $x=A^{-1}y+a$. 由定义 4.1 得

$$f^*(x^*)=\sup_x\{\langle x,x^*\rangle-h[A(x-a)]-\langle x,a^*\rangle-\alpha\}$$

$$=\sup_y\{\langle A^{-1}y+a,x^*\rangle-h(y)-\langle A^{-1}y+a,a^*\rangle-\alpha\}$$

$$=\sup_y\{\langle A^{-1}y,x^*-a^*\rangle-h(y)\}+\langle a,x^*-a^*\rangle-\alpha$$

$$=\sup_y\{\langle y,A^{*-1}(x^*-a^*)\rangle-h(y)\}+\langle a,x^*-a^*\rangle-\alpha$$

$$=h^*[A^{*-1}(x^*-a^*)]+\langle x^*,a\rangle+\alpha^*,$$

其中A^*是A的伴随变换，$\alpha^*=-\alpha-\langle a,a^*\rangle$.

例 9 设偏仿射函数$f(x)=\delta(x\,|\,L+a)+\langle x,a^*\rangle+\alpha$，其中$L$是$R^n$中的子空间，$a,a^*\in R^n$，$\alpha\in R$. 求$f^*(x^*)$.

解 设$h(x)=\delta(x\,|\,L)$，A是恒等线性变换，利用例 7，例 8 的结果可得

$$f^*(x^*)=\delta(x^*-a^*\,|\,L^{\perp})+\langle x^*,a\rangle+\alpha^*$$

$$=\delta(x^*\,|\,L^{\perp}+a^*)+\langle x^*,a\rangle+\alpha^*,$$

其中$\alpha^*=-\alpha-\langle a,a^*\rangle$.

最后简略地叙述一下凹函数的共轭概念．设$g(x)$是R^n中的凹函数，则$g(x)$的共轭$g^*(x^*)$定义为：

$$g^*(x^*)=\inf_x\{\langle x,x^*\rangle-g(x)\}. \tag{26}$$

考虑到$g(x)$是凹函数的等价条件是$-g(x)$是凸函数，故凸函数的共轭函数的结果一般对凹函数的共轭函数也是成立的．特别地，$g^{**}=\mathrm{cl}\,g$，其中闭包运算与凸函数的闭包运算相似．但是如果$f(x)$是凸函数，$f(x)=-g(x)$，则$g^*(x^*)$和$-f^*(x^*)$是不相同的，由计算知

$$f^*(x^*)=\sup_x\{\langle x,x^*\rangle-f(x)\}$$

$$=\sup_x\{-[-\langle x,x^*\rangle-g(x)]\}$$

$$=-\inf_x\{\langle x,-x^*\rangle-g(x)\}=-g^*(-x^*).$$

§ 5 支撑函数

在这一章的§1中，我们已经定义 R^n 中的凸集 C 的支撑函数是

$$\delta^*(x^*|C)=\sup\{\langle x,x^*\rangle|x\in C\}. \tag{1}$$

容易计算

$$\inf\{\langle x,x^*\rangle|x\in C\}=-\delta^*(-x^*|C), \tag{2}$$

所以，支撑函数是与线性函数在凸集 C 上的极值有关的一类函数。

1. 支撑函数的性质

定理 5.1 设 C 是 R^n 中的闭凸集，则 $x\in C$ 的等价条件是对于 $\forall x^*\in R^n$，有

$$\langle x,x^*\rangle\leqslant\delta^*(x^*|C). \tag{3}$$

证明 设 $x\in C$，由(1)式知(3)式成立。

反之，设 x_0 满足(3)，但 $x_0\notin C$。则 $\{x_0\}$ 和 C 可以强分离。故由第一章定理 5.6 知存在 x^* 和 $\varepsilon>0$，对于 $\forall x\in C$，满足

$$\langle x,x^*\rangle\leqslant\langle x_0,x^*\rangle-\varepsilon. \tag{4}$$

将(4)式左边对 $x\in C$ 取上确界，得

$$\delta^*(x^*|C)\leqslant\langle x_0,x^*\rangle-\varepsilon,$$

即 $\delta^*(x^*|C)<\langle x_0,x^*\rangle$，所以 x_0 不满足(3)式，矛盾。所以 $x_0\in C$。▌

推论 5.1.1 设 C 是 R^n 中的凸集，则 $x\in \text{cl}C$ 的等价条件是对于 $\forall x^*\in R^n$，有

$$\langle x,x^*\rangle\leqslant\delta^*(x^*|C). \tag{5}$$

证明留给读者。

定理 5.1 说明，闭凸集 $C\subset R^n$ 可以表示成

$$C=\{x|\langle x,x^*\rangle\leqslant\delta^*(x^*|C),\forall x^*\in R^n\}. \tag{6}$$

所以 C 完全由它的支撑函数确定，有时为了证明两个闭凸集相同，

只要证明它们的支撑函数相同就可以了.

定理 5.2 设C是 R^n 中的凸集，则$C\subset\{x|\langle x, x^*\rangle\leqslant\beta\}$的等价条件是$\delta^*(x^*|C)\leqslant\beta$,其中 $\beta\in R$.

证明 设$C\subset\{x|\langle x,x^*\rangle\leqslant\beta\}$,则$x\in C$时,$\langle x,x^*\rangle\leqslant\beta$,故$\delta^*(x^*|C)=\sup\{\langle x,x^*\rangle|x\in C\}\leqslant\beta$. 反之亦然. ▌

定理 5.3 设C是 R^n 中的紧致凸集,$x^*\in R^n$,$x^*\neq\theta$,则有

1) 在C中存在一点$\bar{x}$,满足

$$\delta^*(x^*|C)=\langle\bar{x},x^*\rangle.$$

2) 超平面$H=\{x|\langle x,x^*\rangle=\delta^*(x^*|C)\}$在$\bar{x}$支撑$C$.

3) 从超平面H到θ的距离等于$\delta^*\left(\frac{x^*}{|x^*|}\middle|C\right)$.

证明 1) $\delta^*(x^*|C)$是线性函数$\langle x,x^*\rangle$在C上的上确界，C是紧致集,故这个上确界在C上的某一点$\bar{x}$达到,即

$$\delta^*(x^*|C)=\langle\bar{x},x^*\rangle.$$

2) 超平面H是半空间$\overline{K}=\{x|\langle x, x^*\rangle\leqslant\delta^*(x^*|C)\}$的边界,由定理 5.2 知

$$C\subset\{x|\langle x,x^*\rangle\leqslant\delta^*(x^*|C)\}.$$

而由 1),有

$$\delta^*(x^*|C)=\langle\bar{x},x^*\rangle,$$

所以H是C的过$\bar{x}$的支撑超平面.

3) 由第一章 §2 中构造超平面的过程知道,x^*是超平面$H=\{x|\langle x,x^*\rangle=\delta^*(x^*|C)\}$的法线. 故存在$\lambda>0$,满足$\lambda x^*\in H$,所以$|\lambda x^*|$是超平面$H$到$\theta$的距离. 因为$\lambda x^*\in H$,故$\langle\lambda x^*,x^*\rangle=\delta^*(x^*|C)$. 于是

$$\delta^*\left(\frac{x^*}{|x^*|}\middle|C\right)=\frac{1}{|x^*|}\delta^*(x^*|C)$$

$$=\frac{1}{|x^*|}\langle\lambda x^*, x^*\rangle=\frac{\lambda}{|x^*|}\langle x^*,x^*\rangle$$

$$=\lambda|x^*|=|\lambda x^*|. \blacksquare$$

读者从定理 5.2 和 定理 5.3 可以 理解，为什么将 $\delta^*(x^*|C)$ 称为支撑函数的理由了.

容易证明，下面的关系是成立的：

$$\delta^*(x^*|C)=\delta^*(x^*|\mathrm{ri}C)=\delta^*(x^*|\mathrm{cl}C). \tag{7}$$

定理 5.4 设 C_1,C_2 是 R^n 中的凸集，则 $\mathrm{cl}C_1\subset\mathrm{cl}C_2$ 的等价条件是 $\delta^*(x^*|C_1)\leqslant\delta^*(x^*|C_2)$.

证明 设 $\mathrm{cl}C_1\subset\mathrm{cl}C_2$. 则对 $\forall x^*\in R^n$，有

$$\begin{aligned}\delta^*(x^*|C_1)&=\sup\{\langle x,x^*\rangle\mid x\in\mathrm{cl}C_1\}\\&\leqslant\sup\{\langle x,x^*\rangle\mid x\in\mathrm{cl}C_2\}\\&=\delta^*(x^*|C_2).\end{aligned}$$

反之，设 $\delta^*(x^*|C_1)\leqslant\delta^*(x^*|C_2)$. 如果 $x\in\mathrm{cl}C_1$，则有

$$\langle x,x^*\rangle\leqslant\delta^*(x^*|C_1)\leqslant\delta^*(x^*|C_2).$$

由推论 5.1.1，知 $x\in\mathrm{cl}C_2$. 所以 $\mathrm{cl}C_1\subset\mathrm{cl}C_2$. $\blacksquare$

例 1 设 B 是 R^n 中的欧氏单位球. 则 $\delta^*(x^*|B)$ 的有效定义域为 R^n. 令 $x^*\neq\theta$，则有

$$\begin{aligned}\delta^*(x^*|B)&=\sup\{\langle x,x^*\rangle\mid x\in B\}\\&=\left\langle\frac{x^*}{|x^*|},x^*\right\rangle=\frac{1}{|x^*|}\langle x^*,x^*\rangle\\&=|x^*|.\end{aligned}$$

因为 $\delta^*(\theta|B)=0$，所以 $\delta^*(x^*|B)=|x^*|$. 见图 9.

例 2 设 T 是 R^2 中顶点在 $(0,\pm1)$，$(\pm1,0)$ 的正方形(图 10). 对于 $x^*=(2,1)$，我们求线性函数 $\langle x,x^*\rangle$ 的最大值，$x\in T$. 容易验证，在顶点 $(1,0)$ 处，该线性函数达到它在 T 上的最大值. 这时，

$$\delta^*(x^*|T)=\langle(1,0),(2,1)\rangle=2.$$

从几何上看，点 $(1,0)$ 是 T 中沿方向 x^* 最远的点，如果取 $x^{*\prime}$

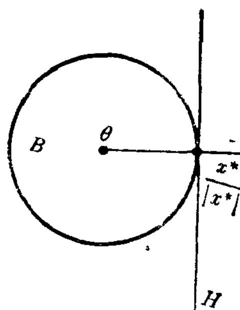


图 9

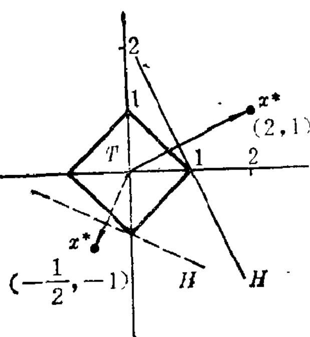


图 10

$=\left(-\frac{1}{2},-1\right)$，则相应的支撑超平面过点$(0,-1)$，且

$$\delta^*(x^{*\prime}|T)=\left\langle(0,-1),\left(-\frac{1}{2},-1\right)\right\rangle=1.$$

因为T是有限的，故$\delta^*(x^*|T)$的有效定义域是R^2。如果设$x^*=(\xi_1^*,\xi_2^*)$，可以算出

$$\delta^*(x^*|T)=\max\{|\xi_1^*|,|\xi_2^*|\}.$$

例3 设在R^2中

$$C=\{x=(\xi_1,\xi_2)\mid 0\leqslant\xi_2\leqslant\xi_1-1\}.$$

因为C是无界集，$\delta^*(x^*|C)$的有效定义域D是R^2中的真子集. 可以算出

$$D=\{(\xi_1^*,\xi_2^*)\mid\xi_1\leqslant 0,\xi_2\leqslant-\xi_1\}.$$

(见图 11). 因为C仅有一个极点$p=(1,0)$，所以其支撑函数为

$$\delta^*(x^*|C)=\langle(1,0),(\xi_1^*,\xi_2^*)\rangle=\xi_1^*,(\xi_1^*,\xi_2^*)\in D.$$

例4 设$C=\{x=(\xi_1,\cdots,\xi_n)\mid\xi_i\geqslant 0,\xi_1+\cdots+\xi_n=1\}$，求$\delta^*(x^*|C)$。这里$C$是凸集，令$x^*=(\xi_1^*,\cdots,\xi_n^*)$，$A=\max\limits_{1\leqslant i\leqslant n}\{\xi_i^*\}$。$\xi_j^*=A$，则$\langle x,x^*\rangle=\xi_1\xi_1^*+\cdots+\xi_n\xi_n^*\leqslant A\sum\limits_{i=1}^{n}\xi_i=A$。而当$i\neq j$时，$\xi_i$

$=0$, $\xi_i=1$ 时，$\langle x, x^*\rangle = A$. 所以

$$\delta^*(x^* | C) = \max_{1 \le i \le n}\{\xi_i^*\}.$$

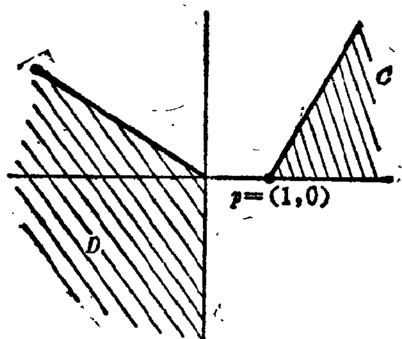


图 11

定理5.5 设C是R^n中的闭凸集，则$\delta(x|C)$和$\delta^*(x^*|C)$是彼此共轭的凸函数.

证明 根据示性函数与支撑函数的定义知

$$[\delta(x|C)]^*(x^*) = \sup\{\langle x, x^*\rangle - \delta(x|C) \mid x \in R^n\}$$
$$= \sup_{x\in C}\langle x, x^*\rangle = \delta^*(x^*|C).$$

从而有

$$[\delta^*(x^*|C)]^*(x) = [\delta(x|C)]^{**}(x) = \mathrm{cl}\,\delta(x|C)$$
$$= \delta(x|\mathrm{cl}\,C) = \delta(x|C). \;\blacksquare$$

定理5.6 设C是R^n中的非空凸集，则$\delta^*(x^*|C)$是闭正齐次正常凸函数，其有效定义域是包含原点的凸锥.

证明 由定理 5.5，$\delta^*(x^*|C)$是闭正常凸函数. 根据支撑函数的定义，在 $0<\lambda<+\infty$时，有

$$\delta^*(\lambda x^*|C) = \sup\{\langle x, \lambda x^*\rangle \mid x \in C\}$$
$$= \lambda \sup\{\langle x, x^*\rangle \mid x \in C\}$$
$$= \lambda\delta^*(x^*|C).$$

故$\delta^*(x^*|C)$是正齐次函数，其有效定义域是锥. 但由定理 1.1，

有效定义域也是凸集，且$\delta^*(\theta|C)=0$，所以有效定义域是包含原点的凸锥. ▌

推论5.6.1 设C是R^n中的非空凸集，则对$\forall x_1^*, x_2^* \in R$，有

$$\delta^*(x_1^*+x_2^*|C) \leqslant \delta^*(x_1^*|C)+\delta^*(x_2^*|C). \tag{8}$$

证明留给读者.

推论5.6.2 设C是R^n中的非空有界闭凸集，则$\delta^*(x^*|C)$是有限正齐次凸函数.

证明 根据定理 5.6，只需说明凸集C有界的必要条件是对于$\forall x^* \in R^n$，$\delta^*(x^*|C)<+\infty$.

事实上，C有界的等价条件是它包含在某个立方体中，从而每一个线性函数在C上有上界，即有

$$\delta^*(x^*|C)<+\infty. \text{ ▌}$$

推论5.6.3 设$f(x)$是R^n上的正齐次凸函数，$f(x) \not\equiv +\infty$，$C=\{x^* | \langle x, x^* \rangle \leqslant f(x), \forall x \in R^n\}$. 则

$$\mathrm{cl} f(x^*)=\delta^*(x^*|C). \tag{9}$$

证明 因为$f(x) \not\equiv +\infty$，故$\mathrm{cl} f(x)$是闭正齐次正常凸函数或常量函数$-\infty$.

如果$f(x) \equiv -\infty$，则$C=\varnothing$，结论明显成立.

如果$f(x)$是正常的，设$\mathrm{cl} f(x^*)=\delta^*(x^*|C)$，现在要求出$C$.

由定理 5.6 及§4的(19)式，有

$$f^*=(\mathrm{cl} f)^*=\delta(\cdot|C),$$

故

$$C=\{x^* | f^*(x^*) \leqslant 0\}. \tag{10}$$

但$f^*(x^*) \leqslant 0$的等价条件是对于$\forall x \in R^n$，$\langle x, x^* \rangle - f(x) \leqslant 0$，所以，有

$$C=\{x^* | \langle x, x^* \rangle \leqslant f(x), \forall x \in R^n\}. \text{ ▌}$$

例5 $f(x)=|x|$是R^n上的正齐次凸函数，由推论 5.6.3，它是一个凸集C的支撑函数. 现在求出这个C.

设$x,y\in R^n$，当$|y|\leqslant 1$时，由Cauchy不等式得
$$|\langle x,y\rangle|\leqslant|x||y|\leqslant|x|,$$
从而$\langle x,y\rangle\leqslant|x|$.

另一方面，当$x=\theta$或$y=\dfrac{x}{|x|}$，即$|y|=1$时，有
$$\langle x,y\rangle=|x|,$$
所以
$$|x|=\sup\{\langle x,y\rangle\,\big|\,|y|\leqslant 1\}=\delta^*(x|B),$$
其中B是欧氏单位球，所以$C=B$.

定理5.7 设$f(x)$是R^n上的闭正齐次凸函数，则
$$f(x)=\delta^*(x|\mathrm{dom}f^*). \tag{11}$$
证明留给读者.

由定理5.5，定理5.6知，如果R^n上的正常凸函数f是示性函数，则它的共轭函数f^*是正齐次的. 而由推论5.6.3，如果f是正齐次的，则它的共轭函数f^*是示性函数. 所以如果f是正齐次的示性函数，则f^*也是正齐次的示性函数. 但正齐次示性函数是凸锥的示性函数，所以如果对于非空凸锥K，$f(x)=\delta(x|K)$，则$f^*(x^*)$也是某一个凸锥$\widetilde{K}$的示性函数，即$f^*(x^*)=\delta(x^*|\widetilde{K})$. 因为$f^*$是闭凸函数，所以$\widetilde{K}$也是闭集.

定义5.1 称$\widetilde{K}$是K的极锥.

由推论5.6.3知
$$\begin{aligned}\widetilde{K}&=\{x^*|\forall x,\langle x,x^*\rangle\leqslant\delta(x|K)\}\\&=\{x^*|\forall x\in K,\langle x,x^*\rangle\leqslant 0\}=-K^*.\end{aligned} \tag{12}$$

定理5.8 设K是R^n中的非空闭凸锥，则$\widetilde{K}$也是非空闭凸锥，且K和$\widetilde{K}$的示性函数彼此共轭，$\widetilde{\widetilde{K}}=\widetilde{(\widetilde{K})}=K$.

证明留给读者.

定理5.8的结论是很重要的，因为在极值问题中常常出现凸锥的示性函数，在确定对偶问题时就需要其共轭函数.

由(12)式知，对于非空闭凸锥K，$\tilde{K}$是K在θ的法线锥；K是$\tilde{K}$在θ的法线锥。

不难验证，当K是R^n中的非负真锥时，$\tilde{K}=-K$。

2. 与规范函数的对偶关系

在§1中，我们已经定义凸集C的规范函数：

$$\gamma(x|C)=\inf\{\lambda>0\,|\,x\in\lambda C\}. \tag{13}$$

为了研究对偶关系，我们还假设$\theta\in\operatorname{int}C$。

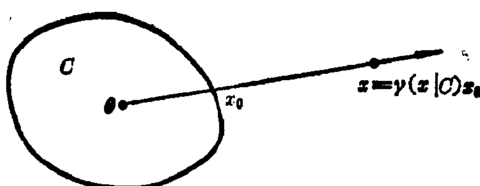


图 12

如图12，设$x\neq\theta$，作从θ出发，通过x的射线，它与C的边界的交点是x_0，则由(13)式，有(严格证明略)

$$x=\gamma(x|C)x_0,$$

$$\gamma(x|C)=|x|/|x_0|. \tag{14}$$

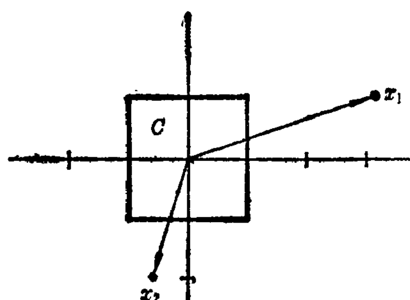


图 13

例6 在R^2中，设C是顶点在$(1,\pm1)$，$(-1,\pm1)$的正方形(图13)。如果从θ出发，通过$x=(\xi_1,\xi_2)$的射线和正方形C的垂直边相交，则根据(14)，有

$$\gamma(x|C)=|\xi_1|.$$

例如$x_1=(3,1)$，则$\gamma(x_1|C)=3$。如果射线与C的水平边相交，则有

$$\gamma(x|C)=|\xi_2|.$$

例如 $x_2=(-\frac{1}{2},-2)$，则$\gamma(x_2|C)=2$。可以验证，对于 $x=(\xi_1,\xi_2)$，有

$$\gamma(x|C)=\max\{|\xi_1|,|\xi_2|\}.$$

如果将例 2 与例 6 比较，不难发现 $\delta^*(x|T)=\gamma(x|C)$，且 $T=C^\circ$. 这个结果不是偶然的巧合，事实上，上面的结论在比较一般的情形下也是成立的.

定理 5.9 设 C 是 R^n 中的紧致凸集，$\theta\in\operatorname{int}C$. 则有

$$\delta^*(x|C)=\gamma(x|C^\circ),\ \delta^*(x|C^\circ)=\gamma(x|C). \tag{14}$$

证明 根据第一章推论 7.3.1 知，C° 是紧致凸集，且 $\theta\in\operatorname{int}C^\circ$.

设 $x\neq\theta$. 如图 14，作从 θ 出发，通过 x 的射线，它与 C° 的边界的交点是 x_0. 因为 $x_0\in C^\circ$，由配极的定义知，对于 $\forall x\in C$，有

$$\langle x,x_0\rangle\leqslant 1.$$

从而

$$\delta^*(x_0|C)=\sup\{\langle x,x_0\rangle|x\in C\}\leqslant 1.$$

又因为 $x=\frac{|x|}{|x_0|}x_0$，故有

$$\delta^*(x|C)=\delta^*\left(\frac{|x|}{|x_0|}x_0\middle|C\right)=\frac{|x|}{|x_0|}\delta^*(x_0|C)$$

$$\leqslant\frac{|x|}{|x_0|}=\gamma(x|C^\circ). \tag{15}$$

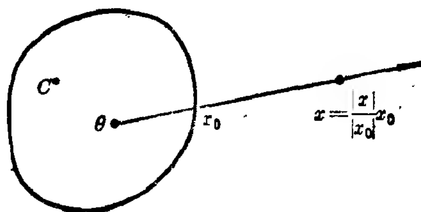


图 14

交换 C 和 C° 的位置，又有

$$\delta^*(x\,|\,C^\circ)\leqslant\gamma(x\,|\,C). \tag{16}$$

另一方面，因为 $x\neq\theta$ 时，$\gamma(x\,|\,C)>0$，故对每一个满足 $0<\alpha<\gamma(x\,|\,C)$ 的 α，有 $x\notin\alpha C$. 但由第一章定理 7.1 和推论 7.3.1，$\alpha C=(\alpha C^{\circ\circ})=\left[\frac{1}{\alpha}C^\circ\right]^\circ$，故 $x\notin\left[\frac{1}{\alpha}C^\circ\right]^\circ$. 这说明在$\frac{1}{\alpha}C^\circ$中存在一点$\frac{1}{\alpha}x^\circ$，满足$\langle\frac{1}{\alpha}x^\circ,x\rangle>1$,即

$$\langle x^\circ,x\rangle>\alpha,\ x^\circ\in C^\circ.$$

所以，有

$$\delta^*(x\,|\,C^\circ)=\sup\{\langle x^*,x\rangle\,|\,x^*\in C^\circ\}>\alpha. \tag{17}$$

因为对每一个 $\alpha<\gamma(x\,|\,C)$,(17)式均成立,于是有

$$\delta^*(x\,|\,C^\circ)\geqslant\gamma(x\,|\,C). \tag{18}$$

交换 C 和 C°的位置,由(18)式又得

$$\delta^*(x\,|\,C)\geqslant\gamma(x\,|\,C^\circ). \tag{19}$$

因为$\delta^*(\theta\,|\,C)=\delta^*(\theta\,|\,C^\circ)=\gamma(\theta\,|\,C)=\gamma(\theta\,|\,C^\circ)=0$，所以由(16)，(18)式得

$$\delta^*(x\,|\,C^\circ)=\gamma(x\,|\,C),$$

由(15)，(19)式得

$$\delta^*(x\,|\,C)=\gamma(x\,|\,C^\circ).\ \blacksquare$$

推论 5.9.1 设 C 是 R^n 中的紧致凸集，$\theta\in\operatorname{int}C$. 则$\delta(x\,|\,C)$与$\gamma(x\,|\,C^\circ)$是彼此共轭的函数,$\delta(x\,|\,C^\circ)$ 与 $\gamma(x\,|\,C)$是彼此共轭的函数.

证明留给读者.

3. 一些函数运算的对偶关系

定理 5.10 设 $h(x)$是 R^n 上的凸函数，则

1)在 $h(x-a)=f(x)$时，$f^*(x^*)=h^*(x^*)+\langle x^*,a\rangle$.

2)在 $h(x)+\langle x,a^*\rangle=f(x)$时，$f^*(x^*)=h^*(x^*-a^*)$.

证明留给读者.

定理 5.11 设 $f(x)$ 是 R^n 上的正常凸函数，则在 $0 \leqslant \lambda < +\infty$ 时，有

$$(\lambda f)^* = f^* \lambda, \tag{20}$$

$$(f\lambda)^* = \lambda f^*. \tag{21}$$

证明 $\lambda > 0$ 时，由共轭函数的定义，有

$$\begin{aligned}(\lambda f)^*(x^*) &= \sup_x \{\langle x, x^*\rangle - \lambda f(x)\} \\ &= \lambda \sup_x \{\langle x, \lambda^{-1} x^*\rangle - f(x)\} \\ &= \lambda f^*(\lambda^{-1} x^*) = (f^* \lambda)(x^*).\end{aligned}$$

$$\begin{aligned}(f\lambda)^*(x^*) &= \sup_x \{\langle x, x^*\rangle - (f\lambda)(x)\} \\ &= \sup_x \{\langle x, x^*\rangle - \lambda f(\lambda^{-1} x)\} \\ &= \lambda \sup_x \{\langle \lambda^{-1} x, x^*\rangle - f(\lambda^{-1} x)\} \\ &= \lambda f^*(x^*).\end{aligned}$$

$\lambda = 0$ 时，因为 $[\delta(x|\theta)]^*(x^*) \equiv 0$，所以定理的结论是明显成立的. ▌

推论 5.11.1 设 C 是 R^n 中的凸集，则在 $0 \leqslant \lambda < +\infty$ 时，有

$$\delta^*(x^* | \lambda C) = \lambda \delta^*(x^* | C). \tag{22}$$

证明 设 $f(x) = \delta(x|C)$，再利用定理. ▌

定理 5.12 设 A 是从 R^n 到 R^m 的线性变换，f 是 R^n 上的凸函数，g 是 R^m 上的凸函数. 则有

$$(Af)^* = f^* A^*, \tag{23}$$

$$((\mathrm{cl}\, g)A)^* = \mathrm{cl}(A^* g^*). \tag{24}$$

如果存在 x，满足 $Ax \in \mathrm{ri}(\mathrm{dom}\, g)$，则有

$$(gA)^* = \mathrm{cl}(A^* g^*). \tag{25}$$

证明 利用共轭函数的定义，有

$$(Af)^*(y^*) = \sup_y \{\langle y, y^*\rangle - \inf_{Ax=y} f(x)\}$$

$$=\sup_{y}\sup_{Ax=y}\{\langle y,y^*\rangle-f(x)\}$$
$$=\sup_{x}\{\langle Ax,y^*\rangle-f(x)\}$$
$$=\sup_{x}\{\langle x,A^*y^*\rangle-f(x)\}$$
$$=f^*(A^*y^*)=(f^*A^*)(y^*),$$

故(23)式成立.

对 A^*,g^* 应用(23)式的结论,有

$$(A^*g^*)^*=g^{**}A^{**}=(\mathrm{cl}\,g)A. \tag{26}$$

所以,对(26)式两边再取一次共轭,有

$$((\mathrm{cl}\,g)A)^*=(A^*g^*)^{**}=\mathrm{cl}(A^*g^*),$$

故(24)式成立.

如果对于一个 $y_0\in\mathrm{ri}(\mathrm{dom}g)$, $g(y_0)=-\infty$, 则由定理 3.5, 对于 $\forall y\in\mathrm{ri}(\mathrm{dom}g)$, $g(y)=-\infty$, 于是 g^*, $(gA)^*$均恒为 $+\infty$. 这时(25)式是显然成立的.

所以设对于 $\forall y,g(y)>-\infty$,且存在 x,满足 $Ax\in\mathrm{ri}(\mathrm{dom}g)$. 由定理 3.17,知$(\mathrm{cl}\,g)A=\mathrm{cl}(gA)$,所以

$$((\mathrm{cl}\,g)A)^*=(gA)^*.$$

由(24)式,最后得到(25)式. ▌

例 7 设 $f(\xi_1,\xi_2)$是 R^2 上的凸函数,令

$$h(\xi_1)=\inf_{\xi_2}f(\xi_1,\xi_2),\xi_1\in R.$$

现在利用定理 5.12,求 $h^*(\xi_1^*)$.

设 A 是投影变换:$(\xi_1,\xi_2)\longrightarrow\xi_1$,则 $h=Af$. 计算知伴随变换 $A^*:\xi_1^*\longrightarrow(\xi_1^*,0)$.

由定理 5.12 中的(23)式知

$$h^*(\xi_1^*)=f^*(\xi_1^*,0).$$

例 8 设 $g_1,\cdots,g_m$ 是 R 上的闭正常凸函数,$a_1,\cdots,a_m\in R^n$, 令

$$h(x)=g_1(\langle a_1,x\rangle)+\cdots+g_m(\langle a_m,x\rangle),$$

这里亦可利用定理 5.12 求出 h^*.

设线性变换 A 是：

$$x\longrightarrow(\langle a_1,x\rangle,\cdots,\langle a_m,x\rangle).$$

令

$$g(y)=g_1(\eta_1)+\cdots+g_m(\eta_m),y=(\eta_1,\cdots,\eta_m),$$

则 $h=gA$. 伴随变换 A^* 是：

$$y^*=(\eta_1^*,\cdots,\eta_m^*)\longrightarrow\eta_1^*a_1+\cdots+\eta_m^*a_m.$$

而

$$g^*(y^*)=g_1^*(\eta_1^*)+\cdots+g_m^*(\eta_m^*),$$

$$(A^*g^*)(x^*)=\inf\{g_1^*(\eta_1^*)+\cdots+g_m^*(\eta_m^*)\mid(\eta_1^*,\cdots,\eta_m^*)\to\eta_1^*a_1+\cdots+\eta_m^*a_m\},$$

所以由定理 5.12 的(25)式,得

$$h^*(x^*)=\mathrm{cl}(A^*g^*)(x^*).$$

定理 5.13 设 $f_1,\cdots,f_m$ 是 R^n 上的正常凸函数,则有

$$(f_1\Box\cdots\Box f_m)^*=f_1^*+\cdots+f_m^*,\tag{27}$$

$$(\mathrm{cl}\,f_1+\cdots+\mathrm{cl}\,f_m)^*=\mathrm{cl}(f_1^*\Box\cdots\Box f_m^*).\tag{28}$$

如果 $\bigcap\limits_{i=1}^{m}\mathrm{ri}(\mathrm{dom}\,f_i)\neq\varnothing$,则(28)式变为

$$(f_1+\cdots+f_m)^*=\mathrm{cl}(f_1^*\Box\cdots\Box f_m^*).\tag{29}$$

证明 利用共轭函数的定义，有

$$\begin{aligned}(f_1\Box\cdots\Box f_m)^*(x^*)&=\sup_x\{\langle x,x^*\rangle-\inf_{x=x_1+\cdots+x_m}[f_1(x_1)+\cdots+f_m(x_m)]\}\\&=\sup_x\sup_{x=x_1+\cdots+x_m}\{\langle x,x^*\rangle-f_1(x_1)-\cdots-f_m(x_m)\}\\&=\sup_{x_1,\cdots,x_m}\{\langle x_1,x^*\rangle+\cdots+\langle x_m,x^*\rangle\end{aligned}$$

$$-f_1(x_1)-\cdots-f_m(x_m)\}$$
$$=f_1^*(x^*)+\cdots+f_m^*(x^*),$$

故(27)式成立.

对 $f_1^*,\cdots,f_m^*$ 应用(27)式得

$$(f_1^*\Box\cdots\Box f_m^*)^*=f_1^{**}+\cdots+f_m^{**}=\mathrm{cl}\, f_1+\cdots+\mathrm{cl}\, f_m. \tag{30}$$

故有

$$(\mathrm{cl}\, f_1+\cdots+\mathrm{cl}\, f_m)^*=(f_1^*\Box\cdots\Box f_m^*)^{**}$$
$$=\mathrm{cl}(f_1^*\Box\cdots\Box f_m^*),$$

所以(28)式成立.

如果 $\bigcap_{i=1}^{m}\mathrm{ri}(\mathrm{dom}\, f_i)\neq\varnothing$，则由定理 3.15，有

$$\mathrm{cl}\, f_1+\cdots+\mathrm{cl}\, f_m=\mathrm{cl}(f_1+\cdots+f_m). \tag{31}$$

所以对(31)式两边取共轭，有

$$(\mathrm{cl}\, f_1+\cdots+\mathrm{cl}\, f_m)^*=[\mathrm{cl}(f_1+\cdots+f_m)]^*$$
$$=(f_1+\cdots+f_m)^*.$$

故(29)式成立. ▌

例 9 设 C 是 R^n 中的非空凸集，令

$$f(x)=d(x,C)=\inf\{|x-y|\mid y\in C\}.$$

下面利用定理 5.13 计算 f^*.

设 $f_1(x)=|x|$，$f_2(x)=\delta(x|C)$，则有

$$f(x)=\inf_y\{|x-y|+\delta(y|C)\}=(f_1\Box f_2)(x).$$

由本节例 1 知，$f_1^*(x^*)=\delta(x^*|B)$ 其中 B 是欧氏单位球. 所以根据定理 5.13 的(27)式，有

$$f^*(x^*)=f_1^*(x^*)+f_2^*(x^*)=\begin{cases}\delta^*(x^*|C) & |x^*|\leqslant 1,\\ +\infty & \text{其他}.\end{cases}$$

例 10 设 $x=(\xi_1,\cdots,\xi_n)$，令 $\|x\|_\infty=\max\{|\xi_i|\mid i=1,\cdots,n\}$. $a_1,\cdots,a_m\in R^n$，令

$$f(x)=\inf\{\|x-\eta_1 a_1-\cdots-\eta_m a_m\|_\infty \mid \eta_i\in R,\ i=1,\cdots,m\}.$$

下面利用定理 5.13 计算 f^*.

设 $f_1(x)=\|x\|_\infty$, $f_2(x)=\delta(x\mid L)$, 其中 $L=\text{span}\{a_1,\cdots,a_m\}$, 故有

$$f(x)=(f_1\square f_2)(x).$$

设 $D=\{x^*=(\xi_1^*,\cdots,\xi_n^*)\mid |\xi_1^*|+\cdots+|\xi_n^*|\leqslant 1\}$, 由本章习题 40 知, $f_1(x)=\delta^*(x\mid D)$, $f_1^*(x^*)=\delta(x^*\mid D)$. 但 $f_2(x)=\delta^*(x\mid L^\perp)$, $f_2^*(x^*)=\delta(x^*\mid L^\perp)$, 其中 $L^\perp$ 是 L 的正交补空间:

$$L^\perp=\{x^*\mid\langle x^*,a_i\rangle=0, i=1,\cdots,m\}.$$

故由(27)式得

$$f^*(x^*)=f_1^*(x^*)+f_2^*(x^*)=\delta(x^*\mid D\cap L^\perp).$$

例 11 设 $h(x)$ 是 R^n 上的闭正常凸函数,令

$$f(x)=\begin{cases}h(x) & x\geqslant\theta,\\ +\infty & \text{其他}.\end{cases}$$

现在利用(28)式求 $f^*(x^*)$.

设 K 是 R^n 中的非负真锥,显然有

$$f(x)=h(x)+\delta(x\mid K).$$

因为 $\tilde{K}=-K$, 故由定理 5.8, 得

$$[\delta(x\mid K)]^*(x^*)=\delta(x^*\mid -K).$$

于是由(28)式, 有

$$\begin{aligned}f^*(x^*)&=\text{cl}[h^*\square\delta(\cdot\mid -K)](x^*)\\&=\text{cl}\inf_{y^*}\{h^*(x^*-y^*)+\delta(y^*\mid -K)\}\\&=\text{cl}\inf\{h^*(z^*)\mid z^*\geqslant x^*\}.\end{aligned}$$

习 题

1. 证明定理 1.4.

2．证明定理1.5．

3．证明下列函数在 R^2 或 R^3 中是凸函数，或者是凹函数，或者两者都不是。

1) $f(x,y,z)=x^2+y^2-4z^2$,

2) $f(x,y)=\begin{cases}(x+y+1)^2 & x+y+1>0,\\ +\infty & 其他,\end{cases}$

3) $f(x,y)=e^{xy}$.

4．设 $f(x,y)=\phi(x^2+y^2)$. 其中 $\phi\in C^{(2)}$ 是增加凹函数．证明 f 在圆 $x^2+y^2\leqslant a^2$ 上是凸函数的充分必要条件是 $0\leqslant u\leqslant a^2$ 时，$\phi'(u)+2u\phi''(u)\geqslant 0$。

5．利用4题的结果求使 f 在 $x^2+y^2\leqslant a^2$ 上是凸函数的最大 a.

1) $f(x,y)=\log(x^2+y^2+1)$,

2) $f(x,y)=\sin(x^2+y^2)$.

6．设 $a\in R^n, \alpha\in R$. 证明 $f(x)=\langle x,a\rangle+\alpha$ 是仿射函数.

7．完成§1例3的证明.

8．设 $x\in R^n$. 证明 $f(x)=|x|$ 是正齐次凸函数.

9．利用凸性证明：

$$(x_1+\cdots+x_m)/m\geqslant(x_1\cdot\cdots\cdot x_m)^{1/m}.$$

其中 $x_1,\cdots,x_m$ 大于0.

10．设 $x,y>0, 0\leqslant t\leqslant 1$. 证明 $tx+(1-t)y\geqslant x^t y^{1-t}$.

11．设 $f(x)$ 在 R^n 中连续，且

$$f\left[\frac{1}{2}(x+y)\right]\leqslant\frac{1}{2}f(x)+\frac{1}{2}f(y).$$

证明 $f(x)$ 是凸函数.

12．证明推论1.8.1.

13．证明推论1.9.1.

14．证明推论1.9.2.

15．证明规范函数是正齐次凸函数.

16．证明定理2.4.

17．证明定理2.9.

18．设 $f(x)$、$g(x)$ 在 x 下半连续，证明 $(f+g)(x)$，$\inf(f,g)$ 在 x 也是下

半连续的.

19. 设$f(x)$、$g(x)$是R^n上的凸函数. $\mathrm{ri}(\mathrm{dom}\, f)=\mathrm{ri}(\mathrm{dom}\, g)=C$. 对于$\forall x\in C$, $f(x)=g(x)$. 证明$\mathrm{cl}\, f(x)=\mathrm{cl}\, g(x)$.

20. 设$f(x)$是R^n上的凸函数, $\mathrm{dom}\, f=C$. 证明$f(x)$在$x_0\in C$下半连续的充分必要条件是对于$\forall \varepsilon>0$,存在$\eta>0$, $\forall x\in C$满足:

1) $|x-x_0|\leqslant\eta$时, $f(x_0)\leqslant f(x)+\varepsilon$.

2) $|x-x_0|>\eta$时, $f(x_0)\leqslant f(x)+\dfrac{\varepsilon}{\eta}|x-x_0|$.

21. 证明推论3.5.2.

22. 证明推论3.6.3.

23. 证明推论3.13.1.

24. 证明在定理4.3中, $P_\varepsilon\cap\mathrm{epi}\, f=\varnothing$.

25. 证明定理4.6.

26. 在R中,设

1) $f(x)=\dfrac{1}{p}|x|^p$, $1<p<+\infty$.

2) $f(x)=\begin{cases}-\dfrac{1}{p}x^p & x\geqslant 0,\\ +\infty & x<0.\end{cases}$ $0<p<1$.

3) $f(x)=\begin{cases}-(a^2-x^2)^{\frac{1}{2}} & |x|\leqslant a,\\ +\infty & |x|>a.\end{cases}$

求$f^*(x^*)$.

27. 如果R^n上的实值函数$f(x)$满足

$$f(x)=f_1(\xi_1)+\cdots+f_n(\xi_n),\ x=(\xi_1,\cdots,\xi_n),$$

则称$f(x)$是可分的. 试证可分凸函数$f(x)$的共轭函数$f^*(x^*)$也是可分的. 设

$$f(x)=\frac{1}{2}(\lambda_1\xi_1^2+\cdots+\lambda_n\xi_n^2),\ 0<\lambda_i<+\infty,$$

计算$f^*(x^*)$以验证上面的结论.

28. 设$f(x)$是R^n上的闭正常凸函数,试证$\inf\limits_x f(x)=0=f(\theta)$的等价条件是$\inf\limits_{x^*} f^*(x^*)=0=f^*(\theta)$.

29. 如果 R^n 中的闭正常凸函数 $f(x)$ 满足 $f(-x)=f(x)$，则称 $f(x)$ 是对称的. 试证 $f(x)$ 对称的等价条件是 $f^*(x^*)$ 对称.

30. 在 R^n 中，设 $x=(\xi_1,\cdots,\xi_n)$，

$$f(x)=\begin{cases}\left(\sum\limits_{i=1}^{n}\xi_i^p\right)^{1/p} & x\geqslant\theta,\\ +\infty & \text{其他}.\end{cases}$$

其中 $p\in R$, $p\neq 0$. 求 $f^*(x^*)$.

31. 设 $f(x)$ 是 R^n 上的凸函数，试证 $\lambda>0$ 时，$f=\lambda f$ 的等价条件是 $f^*=f^*\lambda$.

32. 证明推论 5.1.1.

33. 证明推论 5.6.1.

34. 证明定理 5.7.

35. 证明定理 5.8.

36. 证明推论 5.9.1.

37. 证明定理 5.10.

38. 设 C_1,C_2 是 R^n 中的非空凸集，试证

$$\delta^*(x^*|C_1+C_2)=\delta^*(x^*|C_1)+\delta^*(x^*|C_2).$$

39. 设 B 是 R^n 中的欧氏单位球，$a\in R^n$，$\lambda>0$. 求 $a+\lambda B$ 的支撑函数.

40. 设

$$C_1=\{x=(\xi_1,\cdots,\xi_n)\,|\,|\xi_1|+\cdots+|\xi_n|\leqslant 1\},$$
$$C_2=\{x=(\xi_1,\xi_2)\,|\,\xi_1<0,\xi_2\leqslant\xi_1^{-1}\},$$
$$C_3=\{x=(\xi_1,\xi_2)\,|\,2\xi_1+\xi_2^2\leqslant 0\}.$$

求 C_1,C_2,C_3 的支撑函数.

41. 设 C 是 R^n 中的凸集，$a\in R^n$，试证

$$\delta^*(x^*|C+a)=\delta^*(x^*|C)+\langle a,x^*\rangle.$$

42. 设 C_1,C_2 是 R^n 中的紧致凸集，$\theta\in\text{int }C_1,\theta\in\text{int }C_2$. 试证下面的结论成立.

1) $\delta^*(x^*|C)=\max\{\delta^*(x^*|C_1),\delta^*(x^*|C_2)\}$，其中 $C=\text{co}\{C_1\cup C_2\}$.

2) $\gamma(x|C_1\cap C_2)=\max\{\gamma(x|C_1),\gamma(x|C_2)\}$.

43. 设 C 是 R^n 中的相对开凸集，$f_1(x),f_2(x),\cdots$，是在 C 上有限的凸

函数序列且在 C 的稠密子集 C_1 上逐点收敛. 试证对于 $\forall x\in C$, $f(x)=\lim\limits_{i\to\infty} f_i(x)$ 存在且是有限的, 在 C 的每一个闭有界子集上, $\{f_i(x)\}$ 一致收敛到 $f(x)$.

44. 试证在定理 5.12 的条件下, 有

$$(gA)^*(x^*)=\inf\{g^*(y^*)\mid A^*y^*=x^*\},$$

且对每一个 x^*, 上式中的上确界达到或为 $+\infty$。

45. 试证在定理 5.13 的条件下, 有

$$(f_1+\cdots+f_m)^*(x^*)=\inf\{f_1^*(x_1^*)+\cdots+f_m^*(x_m^*)\mid x^*=x_1^*+\cdots+x_m^*\},$$

且对每一个 x^*, 上式中的下确界达到.

第三章　凸函数的微分

本章讨论凸函数的微分学．凸函数在其有效定义域的内部一定连续，但并不一定在常义下处处可微，所以我们将引入一个新的概念——次微分，从而凸函数在其连续点上总是次可微的．

§1　单边方向导数和次微分

1．单边方向导数

定义 1.1　设 $f(x)$ 是 R^n 到 $[-\infty,+\infty]$ 的实值函数，x 是使 $f(x)$ 为有限的点，如果极限

$$\lim_{\lambda\downarrow 0}\frac{f(x+\lambda y)-f(x)}{\lambda} \tag{1}$$

存在（可以取 $+\infty$ 或 $-\infty$），则称这个极限是 $f(x)$ 在 x 沿向量 y 的单边方向导数，记为 $f'(x:y)$．

注意到

$$-f'(x:y)=\lim_{\lambda\downarrow 0}-\frac{f(x+\lambda y)-f(x)}{\lambda}$$

$$=\lim_{\lambda\uparrow 0}\frac{f(x-\lambda y)-f(x)}{\lambda},$$

故由双边方向导数的概念知 $f'(x:y)$ 是双边方向导数的等价条件是 $f'(x:-y)$ 存在且

$$f'(x:-y)=-f'(x:y). \tag{2}$$

定理 1.1　设 $f(x)$ 是 R^n 上的正常凸函数，$x_0\in\mathrm{dom}f$，则对 $\forall y\in R^n$，$f'(x_0:y)$ 存在．

证明　如果对于 $\forall\lambda>0$，$x_0+\lambda y\notin\mathrm{dom}f$，则 $f(x_0+\lambda y)=+\infty$，$f'(x_0:y)=+\infty$．

如果 λ 充分小时，$x_0+\lambda y\in \mathrm{dom}f$，则由第二章定理3.10，差商$\dfrac{f(x_0+\lambda y)-f(x_0)}{\lambda}$是 λ 的不减函数，所以极限

$$f'(x_0:y)=\lim_{\lambda\downarrow 0}\frac{f(x_0+\lambda y)-f(x_0)}{\lambda}$$

$$=\inf_{\lambda>0}\frac{f(x_0+\lambda y)-f(x_0)}{\lambda}$$

存在. ▎

定理 1.2 设$f(x)$是R^n上的正常凸函数.如果对于$\lambda\in[-\varepsilon,\varepsilon]$，$\varepsilon>0$，$x_0+\lambda y\in \mathrm{dom}f$，则 $f'(x_0:y)$有限.

证明 在 $\alpha_0=-\varepsilon$，$\alpha_1=0$，$\alpha_2=\lambda$ 时应用第二章定理3.10，有

$$\frac{f(x_0)-f(x_0-\varepsilon y)}{\varepsilon}\leqslant\frac{f(x_0+\lambda y)-f(x_0)}{\lambda}. \tag{3}$$

(3)式右边在 $\lambda\downarrow 0$ 时取极限，得

$$\frac{f(x_0+\lambda y)-f(x_0)}{\lambda}\geqslant f'(x_0:y)$$

$$\geqslant\frac{f(x_0)-f(x_0-\varepsilon y)}{\varepsilon}, \tag{4}$$

即 $f'(x_0:y)$有限. ▎

定理 1.3 设 $f(x)$是R^n 上的正常凸函数，则$f'(x:y)$是 y的正齐次凸函数，$f'(x:\theta)=0$，且对$\forall y$，有

$$-f'(x:-y)\leqslant f'(x:y). \tag{5}$$

证明 由定义 1.1，在$\alpha>0$ 时，有

$$f'(x:\alpha y)=\lim_{\lambda\downarrow 0}\frac{f(x+\lambda\alpha y)-f(x)}{\lambda}$$

$$=\alpha\lim_{\lambda\downarrow 0}\frac{f[x+(\lambda\alpha)y]-f(x)}{\lambda\alpha}$$

$$=\alpha f'(x:y).$$

当 $0\leqslant\lambda_1\leqslant 1$ 时，有

$$\frac{f[x+\lambda(\lambda_1 y_1+(1-\lambda_1)y_2)]-f(x)}{\lambda}$$

$$=\frac{f[\lambda_1(x+\lambda y_1)+(1-\lambda_1)(x+\lambda y_2)]-\lambda_1 f(x)-(1-\lambda_1)f(x)}{\lambda}$$

$$\leqslant\lambda_1\frac{f(x+\lambda y_1)-f(x)}{\lambda}+(1-\lambda_1)\frac{f(x+\lambda y_2)-f(x)}{\lambda},$$

在 $\lambda\downarrow 0$ 时对上面的不等式取极限，得

$$f'(x:\lambda_1 y_1+(1-\lambda_1)y_2)$$
$$\leqslant\lambda_1 f'(x:y_1)+(1-\lambda_1)f'(x:y_2).$$

所以 $f'(x:y)$ 是 y 的正齐次凸函数.

显然，$f'(x:\theta)=0$.

任给 $\mu_1>f'(x:-y)$，$\mu_2>f'(x:y)$，因为 $f'(x:y)$ 是 y 的凸函数，根据第二章定理 1.3，有

$$\frac{1}{2}\mu_1+\frac{1}{2}\mu_2>f'\left[x:\frac{1}{2}(-y)+\frac{1}{2}y\right]$$
$$=f'(x:\theta)=0. \tag{6}$$

故对于 $\forall y$，$\frac{1}{2}f'(x:-y)+\frac{1}{2}f'(x:y)\geqslant 0$，即对于 $\forall y$，有

$$f'(x:y)\geqslant -f'(x:-y). \blacksquare$$

容易证明，$f'(x:y)$ 作为 y 的函数，有

$$\operatorname{dom} f'(x:y)=\operatorname{cone}(\operatorname{dom} f-x). \tag{7}$$

当 $f(x)$ 是 R 上的凸函数时，对于左、右导数 $f'_-(x)$，$f'_+(x)$，下面的关系成立：

$$f'_+(x)=\lim_{\lambda\downarrow 0}\frac{f(x+\lambda)-f(x)}{\lambda}=f'(x:1), \tag{8}$$

$$f'_-(x)=\lim_{\lambda\uparrow 0}\frac{f(x+\lambda)-f(x)}{\lambda}=-f'(x:-1). \tag{9}$$

由定理 1.3，得

$$f'_-(x)\leqslant f'_+(x). \tag{10}$$

2. 次梯度与次微分

定义 1.2 设 $f(x)$是R^n上的凸函数. 如果向量 x^*满足:对于 $\forall z$,有

$$f(z) \geqslant f(x) + \langle x^*, z-x \rangle, \tag{11}$$

则称 x^*是 $f(x)$在 x 的次梯度. $f(x)$ 在 x 的次梯度的全体称为 $f(x)$在 x 的次微分,用 $\partial f(x)$ 表示. 如果 $\partial f(x) \neq \varnothing$,则称$f(x)$在 x 是次可微的.

定理 1.4 $\partial f(x)$是闭凸集.

证明留给读者.

由定义 1.2 可知,$f(x)=-\infty$时, $\partial f(x)=R^n$, 而当$f(x)=\infty$,$\mathrm{dom}f \neq \varnothing$ 时, $\partial f(x)=\varnothing$. 为避免上述情形, 以后主要研究正常凸函数.

例 1 设 $f(x)=|x|, x\in R^n$. 求 $\partial f(x)$.

解 当 $x=\theta$ 时,$\partial f(\theta)$中的x^*满足

$$|z| \geqslant \langle x^*, z\rangle, \forall z.$$

因为$|\langle x^*, z\rangle| \leqslant |z||x^*|$.所以 $\partial f(\theta)=B$,其中B是单位球.

$x\neq\theta$ 时,不难验证 $\partial f(x)=|x|^{-1}x$.

我们知道,$f(x)$除 $x=\theta$ 外是可微的, 而由例 1 知 $f(x)$是处处次可微的.

例 2 设

$$f(x)=\begin{cases} -(1-|x|^2)^{\frac{1}{2}} & |x|\leqslant 1, \\ +\infty & |x|>1. \end{cases}$$

可以验证,当$|x|<1$ 时,$f(x)$次可微.而$|x|\geqslant 1$ 时,$\partial f(x)=\varnothing$.

对于例 2 中的 $f(x)$,它仅在 $\mathrm{ri}(\mathrm{dom}f)$处处次可微.

例 3 设C是R^n中的非空凸集. $f(x)=\delta(x|C)$, 则 $x^*\in\partial f(x)$的等价条件是

$$\delta(z|C) \geqslant \delta(x|C) + \langle x^*, z-x\rangle, \forall z. \tag{12}$$

(12)式表示 $x\in C$,$\forall z\in C$ 时,$0\geqslant\langle x^*, z-x\rangle$. 由第一章§8 知 x^*

是C在x的法线向量，所以$\partial\delta(x|C)$是C在x的法线锥.

下面讨论$\partial f(x)$与$f'(x:y)$的关系.

定理 1.5 设$f(x)$是R^n上的凸函数，则x^*是$f(x)$在x的次梯度的充分必要条件是对于$\forall y$，有

$$f'(x:y)\geqslant\langle x^*,y\rangle. \tag{13}$$

证明 设$z=x+\lambda y$，则(11)式变形为

$$[f(x+\lambda y)-f(x)]/\lambda\geqslant\langle x^*,y\rangle,\forall y,\lambda>0. \tag{14}$$

如果x^*是$f(x)$在x的次梯度，则(14)式成立. 在(14)中令$\lambda\downarrow 0$取极限，则对于$\forall y$，有

$$f'(x:y)=\lim_{\lambda\downarrow 0}[f(x+\lambda y)-f(x)]/\lambda\geqslant\langle x^*,y\rangle.$$

反之，如果对于$\forall y$，$f'(x:y)\geqslant\langle x^*,y\rangle$成立，考虑到差商$[f(x+\lambda y)-f(x)]/\lambda$在$\lambda>0$时是$\lambda$的不减函数，故

$$[f(x+\lambda y)-f(x)]/\lambda\geqslant\langle x^*,y\rangle,\forall y,\lambda>0.$$

即(14)式成立，$x^*\in\partial f(x)$. ▌

推论 1.5.1 设$f(x)$是R^n上的凸函数，则

$$\mathrm{cl}f'(x:y)=\delta^*(y|\partial f(x)). \tag{15}$$

证明留给读者.

次梯度不等式(11)的几何意义是：仿射函数$h(z)=f(x)+\langle x^*,z-x\rangle$的图形是凸集$\mathrm{epi}f$在点$(x,f(x))$的不垂直支撑超平面. 故在一维的情形，定理 1.5 说明次梯度是R^2中过$(x,f(x))$而不与$\mathrm{ri}(\mathrm{epi}f)$相交的非垂直直线的斜率$x^*$，它们构成在$f'_-(x)$和$f'_+(x)$之间全体实数组成的闭区间(图 1).

定理 1.6 设$f(x)$是R^n上的凸函数，x是$f(x)$有限的点，则下面的结论成立：

1) 如果$f(x)$在x次可微，则$f(x)$是正常的.

2) 如果$f(x)$在x不是次可微的，则$f(x)$在x一定存在无限的双边方向导数，即一定存在一个y，使

$$f'(x:y)=-f'(x:-y)=-\infty. \tag{16}$$

实际上，对于每一个 $y=z-x, z\in \mathrm{ri}(\mathrm{dom}f)$，(16)式一定成立。

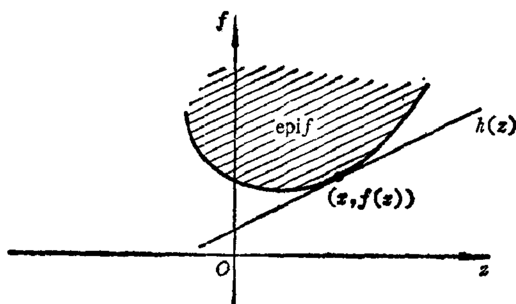


图 1

证明 1) 如果 $f(x)$ 在 x 次可微，由次梯度不等式(11)知，$f(x)$ 是一个仿射函数的优函数。但由条件，$f(x)\neq+\infty$，故 $f(x)$ 是正常的。

2) 如果 $f(x)$ 在 x 不是次可微的，$\partial f(x)=\varnothing$，但 $\partial f(x)=\varnothing$ 的等价条件是 $\delta^*(y\,|\,\partial f(x))=\mathrm{cl}f'(x:y)=-\infty$。于是在某些 y，$f'(x:y)=-\infty$。由定理 1.3，$-f'(x:-y)=-\infty$。即(16)式成立。

设 $\mathrm{dom}f'(x:y)=D$。根据第二章定理 3.5，$y\in\mathrm{ri}D$ 时，$f'(x:y)=-\infty$。令 $C=\mathrm{dom}f-x$。于是 $\theta\in C$。则由(7)式知

$$D=\bigcup_{\lambda>0}\lambda C.$$

于是

$$C\subset D\subset \mathrm{aff}C,$$

故 $\mathrm{ri}C\subset\mathrm{ri}D\subset\mathrm{aff}C$。所以在 $\mathrm{ri}(\mathrm{dom}f)-x$ 上(16)式成立。▌

定理 1.7 设 $f(x)$ 是 R^n 上的正常凸函数，则

1) $x\notin\mathrm{dom}f$ 时，$\partial f(x)=\varnothing$。

2) $x\in\mathrm{ri}(\mathrm{dom}f)$ 时，$\partial f(x)\neq\varnothing$，$f'(x:y)$ 是 y 的闭凸函数，且

$$f'(x:y)=\sup\{\langle x^*,y\rangle\,|\,x^*\in\partial f(x)\}=\delta^*(y\,|\,\partial f(x)). \quad (17)$$

3) $\partial f(x)$ 非空有界的充分必要条件是 $x\in\mathrm{int}(\mathrm{dom}f)$，在这时 $f'(x:y)$ 处处有限，(17)式中的“sup”变为“max”。

证明 1) $x\notin \mathrm{dom}f$ 时，$f(x)=+\infty$。这时只要取 $z\in\mathrm{dom}f$，则(11)式对任何 x^* 都不能满足，故 $\partial f(x)=\varnothing$。

2) 如果 $x\in\mathrm{ri}(\mathrm{dom}f)$. 考虑到 $\mathrm{dom}f'(x:y)$ 是子空间，$f'(x:\theta)=0$，于是 $f'(x:y)$ 在 $\mathrm{dom}f'(x:y)$ 上不恒为 $+\infty$，所以 $f'(x:y)$ 是正常的. 根据第二章推论 3.7.1，$f'(x:y)$ 也是闭的，$\partial f(x)\neq\varnothing$. 由推论 1.5.1，(17)式成立.

3) 根据第二章推论5.6.2，非空闭凸集有界的等价条件是它的支撑函数处处有限，下面将利用这个结论.

如果 $x\in\mathrm{int}(\mathrm{dom}f)$，故 $\mathrm{ri}(\mathrm{dom}f)=\mathrm{int}(\mathrm{dom}f)$. 由2)知，$\mathrm{dom}f'(x:y)$ 是子空间，故 $\mathrm{dom}f'(x:y)=R^n$. 所以 $f'(x:y)$ 处处有限. 而 $\delta^*(y|\partial f(x))=f'(x:y)$，于是 $\partial f(x)$ 有界.

如果 $\partial f(x)$ 非空有界，则 $\delta^*(y|\partial f(x))$ 处处有限，而 $\mathrm{cl}f'(x:y)=\delta^*(y|\partial f(x))$，故 $f'(x:y)$ 处处有限. 由定义 1.1，对 $\forall y$，存在 $\lambda>0$，使 $x+\lambda y\in\mathrm{dom}f$. 所以 $x\in\mathrm{int}(\mathrm{dom}f)$. 在有界闭集上线性函数达到上确界，故(17)中的“sup”由“max”代替. ∎

定理 1.8 设 $f(x)$ 是 R^n 上的正常凸函数，$f(x)$ 在 x 次可微，$f(x)>\inf f(x)$，则 $C=\{z|f(z)\leqslant f(x)\}$ 在 x 的法线锥 $K_0=\mathrm{clcone}[\partial f(x)]$.

证明 由已知条件，$C\neq\varnothing$. 故由法线锥的定义，x^* 是 C 在 x 的法线向量的等价条件是在 $f(z)<f(x)$ 时，有 $\langle z-x,x^*\rangle\leqslant 0$，令 K_0 是 C 在 x_0 的法线锥，则

$$\begin{aligned}K_0&=\{x^*|f(z)<f(x)\text{时},\langle z-x,x^*\rangle\leqslant 0\}\\&=\{x^*|f(z)<f(x)\text{时},\langle\lambda(z-x),x^*\rangle\leqslant 0,\lambda>0\}.\end{aligned}$$

令 $K=\{y|f'(x:y)<0\}=\{y|y=\lambda(z-x),\ f(z)<f(x),\lambda>0\}$，$\tilde{K}$ 是 K 的极锥，则

$$\begin{aligned}\tilde{K}&=\{x^*|\langle x^*,y\rangle\leqslant 0,y\in K\}\\&=\{x^*|\langle x^*,\lambda(z-x)\rangle\leqslant 0,\lambda(z-x)\in K,\lambda>0\}=K_0.\end{aligned}\tag{18}$$

根据第二章定理 3·9，$\mathrm{cl}\{z \mid f(z) < f(x)\} = \{z \mid \mathrm{cl}f(z) \leqslant f(x)\}$，且

$$\begin{aligned}\mathrm{cl}K &= \{y \mid \mathrm{cl}f'(x:y) \leqslant 0\} = \{y \mid \delta^*(y \mid \partial f(x)) \leqslant 0\} \\ &= \{y \mid \langle y, x^* \rangle \leqslant 0, \forall x^* \in \partial f(x)\} \\ &= \{y \mid \langle y, \lambda x^* \rangle \leqslant 0, \forall \lambda x^* \in \lambda \partial f(x), \lambda > 0\} \\ &= \tilde{K}_1.\end{aligned}$$

其中$K_1 = \mathrm{cone}\partial f(x)$. 由(18)及(19)，得

$$K_0 = \tilde{K} = \widetilde{(\mathrm{cl}K)} = (\tilde{\tilde{K}}_1) = \mathrm{cl}K_1. \ \blacksquare$$

3. 次梯度的对偶性

定理 1.9 设$f(x)$是R^n上的正常凸函数，则下面的条件彼此等价：

1) $x^* \in \partial f(x)$.

2) $\langle z, x^* \rangle - f(z)$在$z = x$达到其关于$z$的上确界.

3) $f(x) + f^*(x^*) \leqslant \langle x, x^* \rangle$.

4) $f(x) + f^*(x^*) = \langle x, x^* \rangle$.

如果$f(x) = \mathrm{cl}f(x)$，还可以加入三个条件：

1*) $x \in \partial f^*(x^*)$.

2*) $\langle z^*, x \rangle - f^*(z^*)$在 $z^* = x^*$达到其关于z^*的上确界.

3*) $x^* \in \partial(\mathrm{cl}f(x))$.

证明 利用次梯度不等式(11)，1)可以写成

$$\langle x, x^* \rangle - f(x) \geqslant \langle z, x^* \rangle - f(z), \forall z. \tag{20}$$

(20)正好证明了1)、2)等价.

在(20)式右边对z取上确界，有

$$\langle x, x^* \rangle - f(x) \geqslant f^*(x^*),$$

这正是3)．当然3)成立时，(20)式成立．故2)、3)等价.

由Fenchel 不等式，3)、4)等价.

对偶地，1*)、2*)、3*)等价于

$$f^{**}(x)+f^{*}(x^{*})=\langle x,x^{*}\rangle. \tag{21}$$

而当 $f(x)=\mathrm{cl}f(x)=f^{**}(x)$时,(21)式与4)等价. ▌

推论 1.9.1 设 $f(x)$是 R^n上的闭正常凸函数,则 $\partial f(x)$和 $\partial f^{*}(x^{*})$作为多值映射恰好互为逆映射,即 $x\in\partial f^{*}(x^{*})$的等价条件是 $x^{*}\in\partial f(x)$.

证明 由定理 1.9 的1*)和1)的等价性即得. ▌

推论 1.9.2 设 $f(x)$是 R^n上的正常凸函数,$f(x)$在 x 次可微. 则 $\mathrm{cl}f(x)=f(x)$,$\partial(\mathrm{cl}f(x))=\partial f(x)$.

证明 由 Fenchel 不等式,有

$$f(x)\geqslant\mathrm{cl}f(x)=f^{**}(x)\geqslant\langle x,x^{*}\rangle-f^{*}(x^{*}).$$

如果 $f(x)$在 x 次可微, 则至少存在一个 x^{*} 使 4) 成立, 这时$\langle x,x^{*}\rangle-f^{*}(x^{*})=f(x)$,故 $\mathrm{cl}f(x)=f(x)$.

再由3*)与1)的等价性知,$\partial(\mathrm{cl}f(x))=\partial f(x)$. ▌

例 4 设 $f(x)=\frac{1}{2}|x|^2,x\in R^n$,求 $\partial f(x)$.

解 设 $x=(\xi_1,\cdots,\xi_n)$. $x^{*}=(\xi_1^{*},\cdots,\xi_n^{*})$,容易证明

$$f^{*}(x^{*})=\frac{1}{2}|x^{*}|^2=\frac{1}{2}(\xi_1^{*2}+\cdots+\xi_n^{*2})$$

由 $f(x)$的闭性及定理 1.9,$x^{*}\in\partial f(x)$ 或 $x\in\partial f^{*}(x^{*})$的等价条件是:

$$\langle x,x^{*}\rangle=\sum_{i=1}^{n}\xi_i\xi_i^{*}=\frac{1}{2}|x|^2+\frac{1}{2}|x^{*}|^2. \tag{22}$$

当且仅当 $\xi_i=\xi_i^{*}$时,(22)式成立,所以 $\partial f(x)=x$.

例 5 设 $f(x)=|x|,x\in R^n$,求 $\partial f(x)$.

解 由第二章§5 例1,$f(x)=\delta^{*}(x|B)$,B 是单位球,则

$$f^{*}(x^{*})=\delta(x^{*}|B)=\begin{cases}0 & x^{*}\in B,\\ +\infty & x^{*}\notin B.\end{cases}$$

故由定理 1.9 的4),$x\neq\theta$ 时,$x^{*}\in\partial f(x)$的等价条件是:

$$|x^*|\leqslant 1,\langle x^*,x\rangle=|x|.$$

所以 $\xi_i^*=\xi_i/|x|$. $i=1,\cdots,n,x=(\xi_1,\cdots,\xi_n),x^*=(\xi_1^*,\cdots,\xi_n^*)$,即

$$\partial f(x)=x/|x|.$$

同样,$x=\theta$ 时,$\partial f(\theta)=B$,B 是单位球.

§2 次微分的连续性

定义 2.1 设 $f(x)$是 R^n上的正常凸函数,则称

$$\operatorname{dom}\partial f=\{x|\partial f(x)\neq\varnothing\}$$

为次微分映射 ∂f 的有效定义域,而称

$$\operatorname{rang}\partial f=\bigcup\{\partial f(x)|x\in R^n\}$$

为 ∂f 的值域.

$\operatorname{dom}\partial f$ 不一定是凸集,但由定理1.7,有

$$\operatorname{ri}(\operatorname{dom}f)\subset\operatorname{dom}\partial f\subset\operatorname{dom}f,$$

所以 $\operatorname{dom}\partial f$ 也与凸集差别不大. 由定理 1.9,∂f 的值域是 ∂f^* 的有效定义域,故

$$\operatorname{ri}(\operatorname{dom}f^*)\subset\operatorname{rang}\partial f\subset\operatorname{dom}f^*.$$

下面讨论次微分的连续性.

1. 一维的情形

为方便起见,将左、右导数 $f'_-(x)$ 和 $f'_+(x)$ 推广到 $\operatorname{dom}f$ 区间之外,在 $\operatorname{dom}f$ 右边的点,设 $f'_+=f'_-=+\infty$. 在 $\operatorname{dom}f$ 左边的点,设 $f'_+=f'_-=-\infty$.

定理 2.1 设 $f(x)$是 R 上的闭正常凸函数,则

1) f'_-、f'_+在 R 上是不减函数,在 $\operatorname{int}(\operatorname{dom}f)$上有限.

2) 当 $z_1<x<z_2$时

$$f'_+(z_1)\leqslant f'_-(x)\leqslant f'_+(x)\leqslant f'_-(z_2). \tag{1}$$

3) 对于$\forall x\in R$,有

$$\lim_{z\downarrow x} f'_+(z)=f'_+(x),\quad \lim_{z\downarrow x} f'_-(z)=f'_+(x). \tag{2}$$

$$\lim_{z\uparrow x} f'_+(z)=f'_-(x),\quad \lim_{z\uparrow x} f'_-(z)=f'_-(x). \tag{3}$$

证明 根据§1的(8),(9)式,有

$$f'_+(x)=f'(x:1),f'_-(x)=-f'(x:-1).$$

由定理 1.1,上面两式完全确定且 $f'_-(x)\leqslant f'_+(x)$. 由推广后的定义,$x\notin \text{dom}f$ 时,这个不等式仍成立.

由差商$[f(x+\lambda y)-f(x)]/\lambda$ 的单调性, $f'_+(x)<+\infty$的充分必要条件是x在 cl(dom$f$)右端点的左边,而 $f'_-(x)>-\infty$的充分必要条件是 x 在 cl(dom$f$)左端点的右边. 所以在 int(dom$f$)上,f'_+、f'_-有限.

如果 $y,z\in \text{dom}f,y<z$,则

$$f'_+(y)\leqslant\frac{f(z)-f(y)}{z-y}=\frac{f(y)-f(z)}{y-z}\leqslant f'_-(z). \tag{4}$$

如果 $y,z\notin \text{dom}f,y<z$. 根据推广了的定义, 亦有 $f'_+(y)\leqslant f'_-(z)$. 如果 y,z 中只有一个在 dom f 中, 这时自然有 $f'_+(y)\leqslant f'_-(z)$. 故 (1)成立. 同时也证明了 f'_-、f'_+是不减函数.

由(1),又得

$$f'_+(x)\leqslant\lim_{z\downarrow x} f'_-(z)\leqslant\lim_{z\downarrow x} f'_+(z). \tag{5}$$

为了证明(2),只需再证反向的不等式成立.

由第二章推论3.8.1,$\lim_{z\downarrow x} f(z)=f(x)$. 故存在 $\varepsilon>0$,在$x<y<x+\varepsilon$时,有

$$\frac{f(y)-f(x)}{y-x}=\lim_{z\downarrow x}\frac{f(y)-f(z)}{y-z}\geqslant\lim_{z\downarrow x} f'_+(z).$$

故

$$\lim_{z\downarrow x} f'_+(z)\leqslant \lim_{y\downarrow x}\frac{f(y)-f(x)}{y-x}=f'_+(x). \tag{6}$$

所以(2)成立. 类似可以证明(3). ▌

正如§1 定理 1.5 后面说明的,对于 $\forall x$

$$\partial f(x)=\{x^*\in R \mid f'_-(x)\leqslant x^*\leqslant f'_+(x)\},\tag{7}$$

下面利用(7)式计算函数的次微分.

例 1 在R中,设

$$f(x)=\begin{cases}+\infty & x<-1,\\ 2 & x=-1,\\ x^2 & -1<x\leqslant 0,\\ x & 0<x\leqslant 1,\\ +\infty & x>1.\end{cases}$$

(图 2). 求$\partial f(x)$.

解 $f(x)$是正常凸函数. 利用左、右导数定义得

$$f'_+(x)=\begin{cases}-\infty & x<-1,\\ -\infty & x=-1,\\ 2x & -1<x<0,\\ 1 & 0\leqslant x<1,\\ +\infty & x=1,\\ +\infty & x>1.\end{cases}$$

$$f'_-(x)=\begin{cases}-\infty & x<-1,\\ -\infty & x=-1,\\ 2x & -1<x\leqslant 0,\\ 1 & 0<x\leqslant 1,\\ +\infty & x>1.\end{cases}$$

所以,由(7)式有

$$\partial f(x)=\begin{cases}\varnothing & x\leqslant -1,\\ 2x & -1<x<0,\\ [0,1] & x=0,\\ 1 & 0<x<1,\\ [1,+\infty] & x=1,\\ \varnothing & x>1.\end{cases}$$

$\partial f(x)$的图形是连续无限曲线,见图 3.

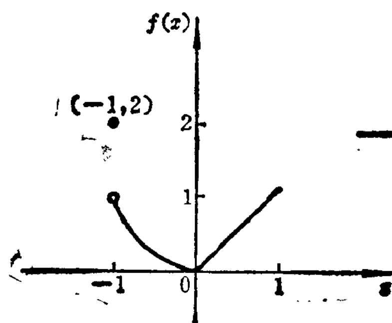


图 2

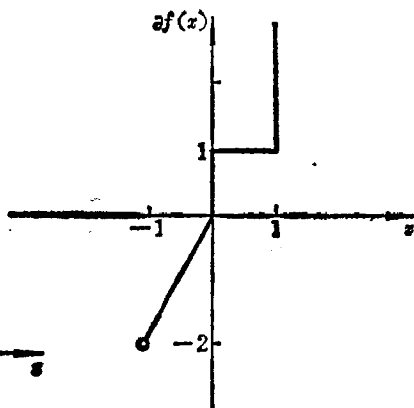


图 3

例 2 在R中，设

$$f(x)=\begin{cases} 0 & x>0, \\ 1 & x=0, \\ +\infty & x<0. \end{cases} \qquad 求\ f'_+(x)。$$

解 计算知

$$f'_+(x)=\begin{cases} 0 & x>0, \\ -\infty & x\leqslant 0. \end{cases}$$

显然，$f'_+(x)$在 $x=0$ 不是右连续的，$f(x)$不是闭的。这个例子说明，定理 2.1 中 $f(x)$是闭的假设是必要的。

当 $f(x)$是闭正常凸函数时，由定理 2.1 中的公式(2)、(3)知，$f'_+(x)$、$f'_-(x)$的每一个都彼此确定另一个。事实上，设 $\varphi(x)$是 R 到$[-\infty,+\infty]$的任意函数，且

$$f'_-(x)\leqslant\varphi(x)\leqslant f'_+(x),\ \forall x\in R.$$

令 $\varphi_+(x)=\lim\limits_{z\downarrow x}\varphi(z)$，$\varphi_-(x)=\lim\limits_{z\uparrow x}\varphi(z)$.

则由定理 2.1，$\varphi(x)$是不减少的，且$f'_+=\varphi_+$，$f'_-=\varphi_-$. 所以 φ 完全确定了 ∂f. 下面的定理说明 φ 是怎样确定 $f(x)$ 到一个常数的.

定理 2.2 设 $a\in R$，$\varphi(x)$是 R 到$[-\infty,+\infty]$且 $\varphi(a)$有限的不减函数，令

$$\varphi_+(x)=\lim_{z\downarrow x}\varphi(z),\quad \varphi_-(x)=\lim_{z\uparrow x}\varphi(z), \tag{8}$$

$$f(x)=\int_a^x\varphi(t)dt, \tag{9}$$

则 $f(x)$是闭正常凸函数，且满足

$$f'_-=\varphi_-\leqslant\varphi\leqslant\varphi_+=f'_+. \tag{10}$$

如果 $g(x)$是另一个闭正常凸函数，满足 $g'_-\leqslant\varphi\leqslant g'_+$，则$g=f+\alpha$，$\alpha$ 是常数.

证明 设 J 是 $\varphi(x)$有限的区间. 因为 $\varphi(x)$ 不减少，$f(x)$作为 Riemann 积分对于 $x\in J$ 是完全确定的，有限的. 在 $\mathrm{cl}J$ 的有限端点，作为Riemann 积分(或 Lebesgue 积分)的极限，$f(x)$是完全确定的. 在 $x\notin\mathrm{cl}\,J$ 时，积分显然是 $+\infty$. 下面证明 $f(x)$在 J 上是凸的，从而由 $\mathrm{cl}J$ 上积分的连续 性得 出 $f(x)$在 R 上是闭的结论.

设 $x,y\in J$，$0<\lambda<1$，$z=(1-\lambda)x+\lambda y$，则

$$\lambda=\frac{z-x}{y-x},\ 1-\lambda=\frac{y-z}{y-x}.$$

而

$$f(z)-f(x)=\int_x^z\varphi(t)dt\leqslant(z-x)\varphi(z), \tag{11}$$

$$f(y)-f(z)=\int_z^y\varphi(t)dt\geqslant(y-z)\varphi(z). \tag{12}$$

故

$$\begin{aligned}&(1-\lambda)[f(z)-f(x)]+\lambda[f(z)-f(y)]\\&\quad\leqslant[(1-\lambda)(z-x)-\lambda(y-z)]\varphi(z)=0,\end{aligned}$$

从而

$$f(z)\leqslant(1-\lambda)f(x)+\lambda f(y), \tag{13}$$

所以 $f(x)$是凸函数.

对于 $\forall x\in J$，在 $\forall z>x$ 时，有

$$\frac{f(z)-f(x)}{z-x}=\frac{1}{z-x}\int_x^z \varphi(t)dt\geqslant\frac{1}{z-x}\varphi(x)(z-x)$$
$$=\varphi(x),$$

故 $f'_+(x)\geqslant\varphi(x)$．相似可证 $x\in J$ 时，$\varphi(x)\geqslant f'_-(x)$．如果 $x\notin J$，这两个不等式显然成立，所以如定理前所述，一定有 $f'_+=\varphi_+, f'_-=\varphi_-$，即 $f'_-=\varphi_-\leqslant\varphi\leqslant\varphi_+=f'_+$.

设 $g(x)$ 是满足 $g'_-\leqslant\varphi\leqslant g'_+$ 的另一个闭正常凸函数，故同样有 $g'_+=\varphi_+, g'_-=\varphi_-$．于是 $g'_+=f'_+, g'_-=f'_-$．由定理2.1中左、右导数的有限性质及 $J\subset\text{dom}f\subset\text{cl}J$，则

$$\text{ri}(\text{dom}g)=\text{ri}(\text{dom}f)=\text{ri}J.$$

因为 $f(x), g(x)$ 是闭的，它们的值由 $\text{ri}J$ 上 $f(x), g(x)$ 的值完全确定．所以只需证明在 $\text{ri}J$ 上

$$g=f+\text{const}. \tag{14}$$

如果 J 仅仅是一个点，(14)式的成立是明显的，故可设 $\text{ri}J=\text{int}J\neq\varnothing$．由定理2.1，在 $\text{int}\,J$ 中 f 和 g 的左、右导数有限，根据极限的可加性，函数 $h=f-g$ 在 $\text{int}J$ 中左、右导数存在，且

$$h'_+(x)=f'_+(x)-g'_+(x)=0$$
$$h'_-(x)=f'_-(x)-g'_-(x)=0.$$

故在 $\text{int}J$ 中 h 的双边导数存在且恒为0．因此在 $\text{int}J$ 中 $f-g=\text{const}$．(14)式成立．▌

为了说明 $\partial f(x)$ 的图形，作下面的定义．

定义 2.2 设 $\varphi(x)$ 是 R 到 $[-\infty,+\infty]$ 的不恒为无限的不减少函数．称 R^2 的子集

$$\Gamma=\{(x,x^*)\mid x\in R, x^*\in R, \varphi_-(x)\leqslant x^*\leqslant\varphi_+(x)\}$$

是完全不减少曲线．

集合 Γ 就象在区间 $I=\{x\mid$ 对于某个 x^*, $(x,x^*)\in\Gamma\}$ 上连续不减少的函数的图形，只是它可能包含垂直线段及水平线段．

实际上它是一条真正的曲线，并且在两端都是无界的．例 1 中的图 3 就是完全不减少曲线的例子．

由定理 2.1、定理 2.2 直接得到：

定理 2.3 R 上定义的闭正常凸函数 $f(x)$ 的次微分映射 ∂f 的图形是 R^2 中的完全不减少曲线 Γ．除一个常数外，Γ 完全确定 $f(x)$．

2．n 维的情形

定理 2.4 设 $f(x)$ 是 R^n 上的闭正常凸函数．如果序列 $\{x_i\}$ 和 $\{x_i^*\}$ 满足 $x_i^*\in\partial f(x_i)$，$\lim\limits_{i\to\infty}x_i=x$，$\lim\limits_{i\to\infty}x_i^*=x^*$，则 $x^*\in\partial f(x)$．换句话说，∂f 的图形是 $R^n\times R^n$ 中的闭子集．

证明 根据定理 1.9，对 $\forall i$ 有

$$\langle x_i,x_i^*\rangle\geqslant f(x_i)+f^*(x_i^*). \tag{15}$$

在 $i\to\infty$ 时取“lim inf”，并利用 $f(x)$，$f^*(x^*)$ 是闭的条件，得

$$\langle x,x^*\rangle\geqslant f(x)+f^*(x^*), \tag{16}$$

即 $x^*\in\partial f(x)$．▮

定理2.5 设 $f(x)$ 是 R^n 上的凸函数，C 是开凸集，$f(x)$ 在 C 中有限，$\{f_i(x)\}$ 是在 C 中有限的凸函数序列，$\lim\limits_{i\to\infty}f_i(x)=f(x)$，$x\in C$．$\{x_i\}\subset C$，$\lim\limits_{i\to\infty}x_i=x$．$\{y_i\}\subset R^n$，$\lim\limits_{i\to\infty}y_i=y$．则

$$\limsup_{i\to\infty} f_i'(x_i:y_i)\leqslant f'(x:y). \tag{17}$$

此外，任给 $\varepsilon>0$，存在一个指标 i_0，在 $i\geqslant i_0$ 时，有

$$\partial f_i(x_i)\subset\partial f(x)+\varepsilon B, \tag{18}$$

其中 B 是单位球．

证明 任给 $\mu>f'(x;y)$，由定义 1.1，存在 $\lambda>0$，$x+\lambda y\in C$，有

$$[f(x+\lambda y)-f(x)]/\lambda<\mu.$$

根据第二章习题 43，知

$$\lim_{i\to\infty}f_i(x_i+\lambda y_i)=f(x+\lambda y),\quad \lim_{i\to\infty}f_i(x_i)=f(x).$$

故 i 充分大时，有

$$[f_i(x_i+\lambda y_i)-f_i(x_i)]/\lambda<\mu. \tag{19}$$

但

$$f_i'(x_i:y_i)\leqslant[f_i(x_i+\lambda y_i)-f_i(x_i)]/\lambda,$$

所以 $\lim\limits_{i\to\infty}\sup f_i'(x_i:y_i)\leqslant\mu$. 由 μ 的任意性，知(17)式成立.

在(17)式中特别取 $y_i=y$，则对 $\forall y\in R^n$，有

$$\lim_{i\to\infty}\sup f_i'(x_i:y)\leqslant f'(x:y). \tag{20}$$

因为 $C\subset\text{int (dom}f)$, $C\subset\text{int(dom}f_i)$，由定理 1.7, $\partial f(x)$, $\partial f_i(x_i)$是有界闭凸集. 于是

$$f_i'(x_i:y)=\delta^*(y\,|\,\partial f_i(x_i)),\ \ f'(x:y)=\delta^*(y\,|\,\partial f(x)),$$

且在 R^n 中处处有限. 于是由第二章习题 43 可知，任给 $\varepsilon>0$，存在 i_0，在 $i\geqslant i_0,\forall y\in B$ 时，有

$$f_i'(x_i:y)\leqslant f'(x:y)+\varepsilon, \tag{21}$$

其中 B 是单位球. 利用方向导数的正齐次性，在 $i\geqslant i_0,\forall y\in R^n$ 时，有

$$f_i'(x_i:y)\leqslant f'(x:y)+\varepsilon|y|, \tag{22}$$

则由定理 1.7 及第二章 § 5 例 1，在 $i\geqslant i_0,\forall y\in R^n$ 时，有

$$\begin{aligned}\delta^*(y\,|\,\partial f_i(x_i))&\leqslant\delta^*(y\,|\,\partial f(x))+\varepsilon\delta^*(y\,|\,B)\\&=\delta^*(y\,|\,\partial f(x)+\varepsilon B).\end{aligned} \tag{23}$$

根据第二章定理 5.4，上面的不等式等价于 $i\geqslant i_0$ 时，

$$\partial f_i(x_i)\subset\partial f(x)+\varepsilon B, \tag{24}$$

故(18)式成立. ▌

§ 3 凸函数的可微性

正如本章开始所指出的，凸函数不一定是处处可微的. 但我们已经证明，在 $\text{ri(dom}f)$ 却一定可以保证正常凸函数的次可微性. 这一节我们专门讨论凸函数的可微性，研究可微性与次可微

的关系，梯度与次梯度的关系.

定义 3.1 设$f(x)$是R^n到$[-\infty,+\infty]$的实值函数，x是$f(x)$有限的点. 如果存在唯一的向量x^*，满足

$$f(z)=f(x)+\langle x^*,z-x\rangle-o(|z-x|), \tag{1}$$

则称$f(x)$在x可微，x^*称为$f(x)$在x的梯度，用$\nabla f(x)$表示.

(1)式可以写成

$$\lim_{z\to x}\frac{f(z)-f(x)-\langle x^*,z-x\rangle}{|z-x|}=0. \tag{2}$$

先求$\nabla f(x)$的表示法. 设$f(x)$在x可微，由(2)，对$\forall y\neq\theta$，$z=x+\lambda y,\lambda>0$时，有

$$\begin{aligned}0&=\lim_{\lambda\downarrow 0}\frac{f(x+\lambda y)-f(x)-\langle\nabla f(x),\lambda y\rangle}{\lambda|y|}\\&=[f'(x:y)-\langle\nabla f(x),y\rangle]/|y|.\end{aligned}$$

故$f'(x:y)$存在且是y的线性函数，对$\forall y$，有

$$f'(x:y)=\langle\nabla f(x),y\rangle. \tag{3}$$

特别地，在$i=1,\cdots,n$时，有

$$\langle\nabla f(x),e_i\rangle=\lim_{\lambda\to 0}\frac{f(x+\lambda e_i)-f(x)}{\lambda}=\frac{\partial f(x)}{\partial\xi_i},$$

其中e_i是R^n中的基本单位向量，$x=(\xi_1,\cdots,\xi_n)$. 所以有

$$\nabla f(x)=\left(\frac{\partial f(x)}{\partial\xi_1},\cdots,\frac{\partial f(x)}{\partial\xi_n}\right). \tag{4}$$

如果设$y=(\eta_1,\cdots,\eta_n)$，则

$$f'(x:y)=\frac{\partial f(x)}{\partial\xi_1}\eta_1+\cdots+\frac{\partial f(x)}{\partial\xi_n}\eta_n. \tag{5}$$

现在专门讨论凸函数的情形.

定理 3.1 设$f(x)$是R^n上的凸函数，$x_0\in R^n$，$x^*\in\partial f(x_0)$，$\nabla f(x_0)$存在，则$x^*=\nabla f(x_0)$.

证明 设$x_k=x_0+\frac{1}{k}(x^*-\nabla f(x_0))$，由$f(x)$在$x_0$的可微性知

$$f(x_k)=f(x_0)+\langle\nabla f(x_0),x_k-x_0\rangle+o(|x_k-x_0|). \tag{6}$$

x^*是$f(x)$在x_0的次梯度,又有

$$-f(x_k)\leqslant -f(x_0)-\langle x^*,x_k-x_0\rangle. \tag{7}$$

(6)+(7)得

$$\begin{aligned}0&\leqslant\langle\nabla f(x)-x^*,x_k-x_0\rangle+o(|x_k-x_0|)\\&=-\frac{1}{k}|\nabla f(x_0)-x^*|^2+\frac{1}{k^2}o(|x^*-\nabla f(x_0)|).\end{aligned} \tag{8}$$

将(8)式两边乘以k,令$k\to\infty$,得

$$-|\nabla f(x_0)-x^*|^2\geqslant 0.$$

所以 $x^*=\nabla f(x_0)$. ▌

定理 3.2 设$f(x)$是R^n上的凸函数,$f(x)$在x有限.

1) 如果$f(x)$在x可微,则$\nabla f(x)$是$f(x)$在x的唯一次梯度,于是对于$\forall z$,有

$$f(z)\geqslant f(x)+\langle\nabla f(x),z-x\rangle. \tag{9}$$

2) 如果$f(x)$在x有唯一次梯度,则$f(x)$在x可微.

证明 1) 这是定理3.1的结果.

2) 设x^*是$f(x)$在x的唯一次梯度,令

$$g(y)=f(x+y)-f(x)-\langle x^*,y\rangle.$$

容易证明$\partial g(\theta)=\{\theta\}$. 由(2)可知,要证明$f(x)$在$x$的可微性,只要证明

$$\lim_{y\to 0}\frac{g(y)}{|y|}=0 \tag{10}$$

即可. 由推论1.5.1知$\mathrm{cl}g'(\theta:y)=\delta^*(y|\partial g(\theta))=0$, 故$g'(\theta$ $y)\equiv 0$. 于是对于$\forall u$,有

$$0=g'(\theta:u)=\lim_{\lambda\downarrow 0}[g(\lambda u)-g(\theta)]/\lambda. \tag{11}$$

注意$g(\theta)=0$,而差商是$\lambda>0$的不减函数,所以凸函数$h_\lambda(u)=$

$g(\lambda u)/\lambda,\lambda>0$,在 $\lambda\downarrow 0$ 时逐点减小到 0. 下面证明在 $u\in B$ 时,这种减小是一致的,其中 B 是单位球.

设 $a_1,\cdots,a_m\in R^n,B\subset\text{co}\{a_1,\cdots,a_m\}$. 故 $u\in B$ 时,有

$$u=\sum_{i=1}^{m}\lambda_i a_i,\ \lambda_i\geqslant 0,\ \sum_{i=1}^{m}\lambda_i=1, \tag{12}$$

且 $0\leqslant h_\lambda(u)\leqslant\sum_{i=1}^{m}\lambda_i h_\lambda(a_i)\leqslant\max\{h_\lambda(a_i)\mid i=1,\cdots,m\}$. 因为对任意 $i,\lambda\downarrow 0$ 时,$h_\lambda(a_i)\downarrow 0$. 故对 $u\in B,\lambda\downarrow 0$ 时,$h_\lambda(u)$一致减小到 0. 换言之,任给 $\varepsilon>0$,存在 $\delta>0$,满足

$$g(\lambda u)/\lambda\leqslant\varepsilon,\forall\lambda\in(0,\delta),\forall u\in B. \tag{13}$$

对于满足 $0<|y|\leqslant\delta$ 的每一个向量 y,可以表示成 $y=\lambda u$,其中 $\lambda=|y|,u\in B$. 故只要 $0<|y|\leqslant\delta$, 就有 $g(y)/|y|\leqslant\varepsilon$,即(10)式成立. ▌

推论 3.2.1 设 $f(x)$是 R^n 上的凸函数. 如果 $f(x)$在 x 有限且可微,则 $f(x)$是正常的凸函数且 $x\in\text{int}(\text{dom}f)$.

证明留给读者.

定理 3.3 设 $f(x)$是R^n上的凸函数,x是使$f(x)$有限的点. 则 $f(x)$在 x 可微的充分必要条件是 $f'(x:y)$是 y 的线性函数或双边偏导数$\frac{\partial f(x)}{\partial\xi_i}$在 x 存在且有限,$i=1,\cdots,n$.

证明 本节开始已经推出,如果 $f(x)$在 x 可微, 则 $f'(x:y)$是 y 的线性函数. 反之,如果 $f'(x:y)$是 y 的线性函数, 则它是闭凸函数且由定理 1.7 知

$$f'(x:y)=\delta^*(y\mid\partial f(x)). \tag{14}$$

故 $\partial f(x)$仅包含一个点. 由定理 3.2 知,$f(x)$在 x 可微.

再证第二个等价条件. 如果 $f(x)$在 x 可微,由本节开始所推证, $\frac{\partial f(x)}{\partial\xi_i}$存在且有限,$i=1,\cdots,n$. 反之,设$\frac{\partial f(x)}{\partial\xi_i}$存在有

限，e_i 是 R^n 的基本单位向量，由§1知

$$f'(x:e_i)=\frac{\partial f(x)}{\partial \xi_i}=-f'(x:-e_i), i=1,\cdots,n.$$

因为 $f'(x:y)$ 是正齐次凸函数，根据第二章定理1.10知 $f'(x:y)$ 在 R^n 上关于 y 是线性的，从而 $f(x)$ 在 x 可微. ▌

定理 3.4 设 $f(x)$ 是在 R 的开区间 I 上有限的凸函数. D 是双边导数 $f'(x)$ 存在的 I 的子集，则 D 包含 I 的最多除去可数个点的所有点（因而 D 在 I 中稠密），且 $f'(x)$ 相对于 D 是连续的，不减少的.

证明 将 $f(x)$ 推广成 R 上的闭正常凸函数. 由定理2.1，$f'_+(x)=f'_-(x)$ 的充分必要条件是 $f'_+(x)$ 在 x 连续，所以 D 由 f'_+ 连续的点组成. I 中不属于 D 的点是不减少函数 f'_+ 具有跳跃的点，而且最多只有可数个这样的跳跃（见[39]）. 因为在 D 上 f'_+ 与 f' 相同，故 $f'(x)$ 相对于 D 是连续的、不减少的. ▌

定理 3.5 设 $f(x)$ 是 R^n 上的正常凸函数，$y\neq\theta$. 令

$$D=\{x\in \text{int}(\text{dom}f)\,|\,f'(x:y)=-f'(x:-y)\}. \tag{15}$$

则 $f'(x:y)$ 作为 x 的函数在 D 中连续，D 在 $\text{int}(\text{dom}f)$ 中处处稠密.

证明 根据定理2.5，$f'(x:y)$ 作为 x 的函数在 $\text{int}(\text{dom}f)$ 中上半连续. 故只要证明 $f'(x:y)$ 在 D 中下半连续即可，即

$$\lim_{z\to x}\inf f'(z:y)=-f'(x:-y),\quad \forall x\in \text{int}(\text{dom}f). \tag{16}$$

首先，“$\geqslant$”一定成立. 因为由定理1.3，有

$$f'(z:y)\geqslant -f'(z:-y), \tag{17}$$

而 $f'(z:-y)$ 在 $\text{int}(\text{dom}f)$ 中关于 z 上半连续.

设 $g(\lambda)=f(x+\lambda y)$，则有

$$\begin{aligned}\lim_{z\to x}\inf f'(x:y)&\leqslant \lim_{\lambda\uparrow 0} f'(x+\lambda y:y)=\lim_{\lambda\uparrow 0} g'_+(\lambda)\\&=g'_-(0)=-f'(x:-y).\end{aligned} \tag{18}$$

即“$\leqslant$”成立，所以(16)式成立.

关于 D 在$\text{int}(\text{dom} f)$中处处稠密的证明略，见[29]. ▌

定理 3.6 设 $f(x)$是 R^n 上的正常凸函数. D 是 $f(x)$可微的点集. 则 D 是 $\text{int}(\text{dom} f)$的稠密子集，即 $\text{int}(\text{dom} f)\backslash D=\bar{D}$ 的测度是 0，梯度映射$\nabla f: x\to\nabla f(x)$在 D 上连续.

证明 设 $e_1,\cdots,e_n$ 是基本单位向量，在 $y=e_i$ 时，应用定理 3.5，则$\dfrac{\partial f(x)}{\partial\xi_i}$存在的子集 $D_i\subset\text{int}(\text{dom} f)$的补 $\bar{D}_i$ 是$\text{int}(\text{dom} f)$ 中的零测度集，而$\overline{\bigcap_{i=1}^{n} D_i}=\bigcup_{i=1}^{n}\bar{D}_i$. 根据定理3.3，$\bigcap_{i=1}^{n} D_i=D$，故 $\bar{D}$的测度为 0，D 在 $\text{int}(\text{dom} f)$中处处稠密.

由定理 3.5，$\dfrac{\partial f}{\partial\xi_i}$在 D_i 上连续，$i=1,\cdots,n$. 故 n 个偏导数在 D 上连续. 而 $\nabla f(x)=\left(\dfrac{\partial f(x)}{\partial\xi_1},\cdots,\dfrac{\partial f(x)}{\partial\xi_n}\right)$. 所以$\nabla f(x)$ 在 D 上连续. ▌

§ 4 一些函数的次微分

1. 凸函数之和的次微分

定理 4.1 设 $f_1(x)$、$f_2(x)$是 R^n 上的正常凸函数，令 $f(x)=f_1(x)+f_2(x)$，则

1) $\partial f(x)\supset\partial f_1(x)+\partial f_2(x),\quad\forall x.$ (1)

2) 如果存在 $x_1\in\text{dom} f_1\cap\text{dom} f_2$，$f_1(x)$在 x_1 连续，则当 $x_0\in\text{dom} f_1\cap\text{dom} f_2$ 时，有

$$\partial f(x_0)=\partial f_1(x_0)+\partial f_2(x_0). \tag{2}$$

证明 1) 设 $x^*=x_1^*+x_2^*$，其中 $x_i^*\in\partial f_i(x)$，$i=1.2$，则对$\forall z$，有

$$f(z)=f_1(z)+f_2(z)$$

$$\geqslant f_1(x)+\langle z-x,x_1^*\rangle+f_2(x)+\langle z-x,x_2^*\rangle$$
$$=f(x)+\langle z-x,x_1^*+x_2^*\rangle$$
$$=f(x)+\langle z-x,x^*\rangle.$$

故 $x^*\in\partial f(x)$，从而 $\partial f(x)\supset\partial f_1(x)+\partial f_2(x)$.

2) 只需证明 $\partial f(x_0)\subset\partial f_1(x_0)+\partial f_2(x_0)$. 为简单计，不妨设 $x_0=\theta,f_1(\theta)=f_2(\theta)=f(\theta)=0$.

设 $x^*\in\partial f(\theta)$，即对 $\forall x$，有

$$f_1(x)+f_2(x)\geqslant\langle x,x^*\rangle. \tag{3}$$

将(3)式改写成对 $\forall x$，有

$$f_1(x)-\langle x,x^*\rangle\geqslant-f_2(x). \tag{4}$$

在 R^{n+1} 中设

$$A=\{(x,\mu)\mid\mu>f_1(x)-\langle x,x^*\rangle\}, \tag{5}$$

$$C=\{(y,\nu)\mid\nu<-f_2(y)\}. \tag{6}$$

因为 $f_1(x),f_2(x)$ 是凸函数，所以 A,C 是 R^{n+1} 中的凸集.

由(4) 式，$A\cap C=\varnothing$，故 A,C 可以分离，存在非零向量 (x_0^*,μ^*)，满足

$$\langle y,x_0^*\rangle+\nu\mu^*\leqslant\langle x,x_0^*\rangle+\mu\mu^*, \tag{7}$$
$$(y,\nu)\in C,(x,\mu)\in A.$$

因为 μ 可以无限增大，由(7) 式知 $\mu^*\geqslant0$. 下面证明 $\mu^*>0$.

设其相反，$\mu^*=0$. 则(7) 式变成

$$\langle y,x_0^*\rangle\leqslant\langle x,x_0^*\rangle, x\in\mathrm{dom}f_1,y\in\mathrm{dom}f_2. \tag{8}$$

设 $y=x_1$，因为 $x_1\in\mathrm{dom}f_1$，$f_1(x)$ 在 x_1 连续，则当 $\varepsilon>0$ 充分小时，$x_1+\varepsilon z\in\mathrm{dom}f_1,z\in B$，$B$ 是单位球. 在(8) 式中代入 $y=x_1$，$x=x_1+\varepsilon z$，得

$$0\leqslant\varepsilon\langle z,x_0^*\rangle,z\in B.$$

故 $x_0^*=\theta$. 这与 (x_0^*,μ^*) 是非零向量矛盾.

故设 $\mu^*=1$. 根据 A、C 的定义，由(7) 式得

$$-f_2(y)+\langle y,x_0^*\rangle\leqslant f_1(x)-\langle x,x^*\rangle+\langle x,x_0^*\rangle, \tag{9}$$

$$x \in \mathrm{dom} f_1, y \in \mathrm{dom} f_2.$$

因为在 $\mathrm{dom} f_1, \mathrm{dom} f_2$之外，$f_1, f_2$均是$+\infty$，故对于$x, y \in R^n$，(9)式仍成立。设 $y = x_0 = \theta$，由(9) 式，有

$$\langle x, x^* - x_0^* \rangle \leqslant f_1(x), \forall x, \tag{10}$$

于是$x^* - x_0^* \in \partial f_1(\theta)$，设$x = x_0 = \theta$，由(9) 式，有

$$\langle y, x_0^* \rangle \leqslant f_2(y), \forall y, \tag{11}$$

于是 $x_0^* \in \partial f_2(\theta)$，则

$$x^* = (x^* - x_0^*) + x_0^* \in \partial f_1(\theta) + \partial f_2(\theta).$$

最后得

$$\partial f(\theta) \subset \partial f_1(\theta) + \partial f_2(\theta). \blacksquare$$

定理 4.2 设 $f_1(x)$，$f_2(x)$ 是 R^n 上的正常凸函数，且 $\mathrm{ri}(\mathrm{dom} f_1) \cap \mathrm{ri}(\mathrm{dom} f_2) \neq \varnothing$，令 $f(x) = f_1(x) + f_2(x)$，则

$$\partial f(x) = \partial f_1(x) + \partial f_2(x), \forall x. \tag{12}$$

证明 由定理4.1的1) 知，只要证明

$$\partial f(x) \subset \partial f_1(x) + \partial f_2(x), \forall x. \tag{13}$$

即可

由第二章定理5.13及习题45，$f^*(x^*)$可以表示为

$$f^*(x^*) = \inf\{f_1^*(x_1^*) + f_2^*(x_2^*) \mid x^* = x_1^* + x_2^*\}, \tag{14}$$

且对于每一个x^*，(14)式中的下确界均达到。

设 $x^* \in \partial f(x)$，由定理1.9的1)、4)，$x^* \in \partial f(x)$的充分必要条件是

$$\langle x, x^* \rangle = f_1(x) + f_2(x) + \inf\{f_1^*(x_1^*) + f_2^*(x_2^*) \mid x^* = x_1^* + x_2^*\},$$

其中对每一个x^*，由某些 x_1^*, x_2^*达到下确界。故$\partial f(x)$由满足

$$\langle x, x_1^* \rangle + \langle x, x_2^* \rangle = f_1(x) + f_2(x) + f_1^*(x_1^*) + f_2^*(x_2^*) \tag{15}$$

的向量 $x_1^* + x_2^*$组成。但由 Fenchel 不等式，总有

$$\langle x, x_i^* \rangle \leqslant f_i(x) + f_i^*(x_i^*), i = 1, 2.$$

而上式等号成立的充要条件是 $x_i^* \in \partial f_i(x)$，故由(15)式，有

$$\partial f(x) \subset \partial f_1(x) + \partial f_2(x), \forall x. \blacksquare$$

推论 4.2.1 设 C_1、C_2 是 R^n 中的凸集，$\text{ri}C_1 \cap \text{ri}C_2 \neq \varnothing$，$K_1$，$K_2$ 是 C_1, C_2 在 x 的法线锥，则 $K_1 + K_2$ 是 $C_1 \cap C_2$ 在 x 的法线锥.

证明留给读者.

推论 4.2.2 设 K_1, K_2 是 R^n 中包含原点在内的凸锥，$\text{ri}K_1 \cap \text{ri}K_2 \neq \varnothing$，则

$$(K_1 \cap K_2)^* = K_1^* + K_2^*. \tag{16}$$

证明 易知 $\delta(x | K_1 \cap K_2) = \delta(x | K_1) + \delta(x | K_2)$, (17)

而 $\text{dom}\delta\ (x | K_i) = K_i$，$i = 1$、$2$. 于是

$\text{ri}\ (\text{dom}\delta\ (x | K_1)\ ) \cap \text{ri}(\text{dom}\delta(x | K_2)) \neq \varnothing$. 对 (17) 式应用定理4.2，并注意到 $\partial\delta(\theta | K_i) = -K_i^*$，得

$$\begin{aligned} -(K_1 \cap K_2)^* &= \partial\delta(\theta | K_1 \cap K_2) \\ &= \partial\delta(\theta | K_1) + \partial\delta(\theta | K_2) = -K_1^* - K_2^*. \end{aligned}$$

即(16)式成立. $\blacksquare$

2. 凸函数在线性变换中的逆象的次微分

定理 4.3 设 $g(y)$ 是 R^m 上的正常凸函数，A 是 R^n 到 R^m 的线性变换，$f(x) = (gA)(x)$，则

1) $\partial f(x) \supset A^* \partial(gA)(x), \forall x.$ (18)

2) 如果 $\text{rang}A \cap \text{ri}(\text{dom}g) \neq \varnothing$，还有

$$\partial f(x) = A^* \partial(gA)(x), \forall x. \tag{19}$$

其中 A^* 是 A 的伴随变换，$\text{rang}A$ 是 A 的值域.

证明 1) 设 $x^* \in A^* \partial(gA)(x)$，则存在一个 $y^* \in \partial(gA)(x)$，满足 $x^* = A^* y^*$. 故对 $\forall z \in R^n$，有

$$\begin{aligned} f(z) = g(Az) &\geq g(Ax) + \langle y^*, Az - Ax \rangle \\ &= f(x) + \langle x^*, z - x \rangle. \end{aligned}$$

所以 $x^* \in \partial f(x)$，(18)式成立.

2) 由1)，只要证明 $\partial f(x) \subset A^* \partial(gA)(x)$ 即可.

因为 $\text{rang}A \cap \text{ri}(\text{dom}g) \neq \varnothing$，则 $f(x)$ 是正常的. 由第二章定

理5.12及习题 44，有

$$f^*(x^*)=\inf\{g^*(y^*)\mid A^*y^*=x^*\}, \tag{20}$$

其中下确界对使 $f^*(x^*)\neq+\infty$的每一个 x^*,在某一个 y^*达到.

设 $x^*\in\partial f(x)$,由定理1.9，有

$$f(x)+f^*(x^*)=\langle x,x^*\rangle. \tag{21}$$

故由(20)式，存在 y^* 满足 $A^*y^*=x^*$, $f(x)+g^*(y^*)=\langle x, A^*y^*\rangle$,即

$$(gA)(x)+g^*(y^*)=\langle Ax,y^*\rangle. \tag{22}$$

再根据定理1.9, $y^*\in\partial g(Ax)$, $x^*\in A^*\partial g(Ax)$.所以

$$\partial f(x)\subset A^*\partial(gA)(x). \blacksquare$$

3. 凸函数族逐点上确界的次微分

定理 4.4 设 $C\subset R^n$ 是闭凸集, $f(x)=\delta^*(x\mid C)$，则

$$\partial f(x)=\{x^*\mid x^*\in C,\langle x,x^*\rangle=f(x)\}. \tag{23}$$

特别地, $\partial f(\theta)=C$.

证明 设 $x^*\in C,\langle x,x^*\rangle=f(x)$.则对$\forall z$,有

$$f(z)-f(x)\geqslant\langle z,x^*\rangle-\langle x,x^*\rangle=\langle z-x,x^*\rangle,$$

即$x^*\in\partial f(x)$,从而$\partial f(x)\supset\{x^*\mid x^*\in C,\langle x,x^*\rangle=f(x)\}$.

反之,设 $x^*\in\partial f(x)$. 先证$x^*\in C$. 设其相反, $x^*\notin C$,则x^*, C 可以强分离. 由第一章习题76知,存在向量 p,满足

$$\sup_{x'}\{\langle p,x'\rangle\mid x'\in C\}<\langle p,x^*\rangle. \tag{24}$$

考虑到 $\sup(M-N)\geqslant\sup M-\sup N$,则有

$$\sup_{x'}\{\langle z-x,x'\rangle\mid x'\in C\}\geqslant f(z)-f(x)\geqslant\langle z-x,x^*\rangle. \tag{25}$$

在(25)式中设 $z=x+p$,得

$$\sup_{x'}\{\langle p,x'\rangle\mid x'\in C\}\geqslant\langle p,x^*\rangle, \tag{26}$$

这与(24)式矛盾,所以 $x^*\in C$.

再证$\langle x,x^*\rangle=f(x)$. 由$x^*\in\partial f(x)$,则对 $\forall z$,有

$$f(z)-\langle z,x^*\rangle \geqslant f(x)-\langle x,x^*\rangle.$$

令 $z=\theta$，得$\langle x,x^*\rangle \geqslant f(x)$．而 $x^*\in C$，又有$\langle x,x^*\rangle \leqslant f(x)$.
所以$\langle x,x^*\rangle = f(x)$，则$\partial f(x)\subset\{x^*\mid x^*\in C,\ \langle x,x^*\rangle = f(x)\}$.
综上述，得到

$$\partial f(x)=\{x^*\mid x^*\in C,\langle x,x^*\rangle = f(x)\}.$$

特别地，令 $x=\theta$，则对于 $\forall x^*\in C$，有 $f(\theta)=\langle\theta,x^*\rangle=0$，所以$\partial f(\theta)=C$. ▌

推论 4.4.1 设$f(x)$是R^n上的正齐次闭凸函数，则

$$\partial f(x)=\{x^*\in \mathrm{dom} f^*\mid\langle x,x^*\rangle = f(x)\}. \tag{27}$$

证明留给读者.

现在讨论一般凸函数族的逐点上确界的次微分.

以下设 I 是指标集，$f(x,\alpha)$是 x 的凸函数，$\alpha\in I$.

$$f(x)=\sup_{\alpha}\{f(x,\alpha)\mid \alpha\in I\}. \tag{28}$$

先证明两个辅助定理.

定理 4.5 设 I 是紧致集，函数 $f(x,\alpha)$ 在 x_0 和 $\alpha\in I$ 的邻域内关于 x 和 α 连续，则函数 $f(x)$在 x_0 连续.

证明 设 $I(x)=\{\alpha\in I\mid f(x,\alpha)=f(x)\}$，则当 $\alpha\in I(x)$，$\alpha_0\in I(x_0)$时，有

$$f(x,\alpha)=f(x)\geqslant f(x,\alpha_0), \tag{29}$$

$$f(x_0,\alpha)\leqslant f(x_0)=f(x_0,\alpha_0). \tag{30}$$

(29)—(30)，得

$$f(x,\alpha)-f(x_0,\alpha)\geqslant f(x)-f(x_0)\geqslant f(x,\alpha_0)-f(x_0,\alpha_0). \tag{31}$$

显然 $x\to x_0$ 时，(31)式的右边趋于0．下面证明左边也趋于0.

设相反，存在序列$\{x_k\}$，$\lim\limits_{k\to\infty}x_k=x_0$，$\alpha_k\in I(x_k)\subset I$，使

$$|f(x_k,\alpha_k)-f(x_0,\alpha_k)|\geqslant\varepsilon>0.$$

因为 I 是紧致集，可以设 $\alpha_k\xrightarrow[k\to\infty]{}\alpha\in I$，故有

$$f(x_k,\alpha_k)-f(x_0,\alpha_k)\xrightarrow[k\to\infty]{}f(x_0,\alpha)-f(x_0,\alpha)=0,$$

这与假设矛盾.

综上述，$x\to x_0$ 时，(31)式左、右两边趋于0，所以 $f(x)\to f(x_0)$，$f(x)$ 在 x_0 连续. ▌

定理 4.6 设定理4.5的条件成立，C 是开集，满足：$I(x_0)\subset C\subset I$，则存在 x_0 的一个邻域，在这个邻域的任何点 x，$I(x)\subset C$.

证明 假设对于满足 $I(x_0)\subset C_0$ 的某个开集 C_0，存在序列 $\{x_k\}$，$\lim\limits_{k\to\infty}x_k=x_0$，并存在 $\alpha_k\in I(x_k)$，$\alpha_k\notin C_0$.

由定义，$f(x_k)=f(x_k,\alpha_k)$，在 $x_k\to x_0$ 时取极限，根据定理4.5得

$$f(x_0)=f(x_0,\alpha_0)$$

即 $\alpha_0\in I(x_0)\subset C_0$，则当 k 充分大时，α_k 属于 α_0 的邻域 C_0，这与 $\alpha_k\notin C_0$ 矛盾. ▌

定理 4.7 设 I 是紧致集，对于 $\alpha\in I$，$f(x,\alpha)$ 是 R^n 上的凸函数. 对于 $\lambda\in[-\delta,\delta]$，$\delta>0$，及 $\alpha\in I$，$f(x_0+\lambda p,\alpha)$ 关于 λ 和 α 连续. $f(x)$ 由(28)式定义，则 $f(x)$ 在 x_0 沿方向 p 的方向导数存在且

$$f'(x_0:p)=\max_{\alpha}\{f'(x_0:p,\alpha)\mid\alpha\in I(x_0)\},\tag{32}$$

其中 $f'(x_0:p,\alpha)$ 是 $f(x,\alpha)$ 在 x_0 沿方向 p 的方向导数.

证明 设

$$g(\lambda)=f(x_0+\lambda p),\quad g(\lambda,\alpha)=f(x_0+\lambda p,\alpha),$$
$$I(\lambda)=I(x_0+\lambda p).$$

显然 $f'(x_0:p)=g'(0)$，$f'(x_0:p,\alpha)=g'(0,\alpha)$.

其中 $g'(0)=\lim\limits_{\lambda\downarrow 0}\dfrac{g(\lambda)-g(0)}{\lambda}$，

$$g'(0,\alpha)=\lim_{\lambda\downarrow 0}\frac{g(\lambda,\alpha)-g(0,\alpha)}{\lambda}.$$

于是(32)式变为

$$g'(0)=\max_{\alpha}\{g'(0,\alpha)\mid\alpha\in I(0)\}.\tag{33}$$

因为$g(\lambda,\alpha)$关于α连续，I是紧致集，故函数

$$g(\lambda)=\max_{\alpha}\{g(\lambda,\alpha)\mid\alpha\in I\}$$

在$\lambda\in[-\delta,\delta]$是有限的，$g'(0)$存在且有限．在(31)式中设$x=x_0+\lambda p,\lambda>0$，得

$$\frac{g(\lambda,\alpha)-g(0,\alpha)}{\lambda}\geqslant\frac{g(\lambda)-g(0)}{\lambda}\geqslant\frac{g(\lambda,\alpha_0)-g(0,\alpha_0)}{\lambda},\tag{34}$$

$$\alpha\in I(\lambda),\alpha_0\in I(0).$$

由g的凸性，(34)式右边的差商在$\lambda\downarrow0$时单调减小，则

$$\frac{g(\lambda)-g(0)}{\lambda}\geqslant g'(0,\alpha_0),\qquad\alpha_0\in I(0).$$

故有

$$g'(0)\geqslant\sup_{\alpha}\{g'(0,\alpha)\mid\alpha\in I(0)\}.\tag{35}$$

再研究(34)式左边的不等式．由定理4.6，对于满足$I(0)=I(x_0)\subset C$的开集C，可以找到$\varepsilon>0$，在$0\leqslant\lambda<\varepsilon$时$I(\lambda)\subset C$．故在$\lambda<\varepsilon$时，由(34)式有

$$g'(0)\leqslant\frac{g(\lambda)-g(0)}{\lambda}\leqslant\sup_{\alpha}\left\{\frac{g(\lambda,\alpha)-g(0,\alpha)}{\lambda}\,\middle|\,\alpha\in C\right\}.\tag{36}$$

在全体$C\supset I(0)$上取(36)式右边的下确界．因为$g(\lambda,\alpha)$关于α连续，$I(0)$是紧致集的闭子集，则不难证明

$$\inf_{C\supset I(0)}\ \sup_{\alpha}\left\{\frac{g(\lambda,\alpha)-g(0,\alpha)}{\lambda}\,\middle|\,\alpha\in C\right\}$$
$$=\max_{\alpha}\left\{\frac{g(\lambda,\alpha)-g(0,\alpha)}{\lambda}\,\middle|\,\alpha\in I(0)\right\}.$$

所以

$$g'(0)\leqslant\max_{\alpha}\left\{\frac{g(\lambda,\alpha)-g(0,\alpha)}{\lambda}\,\middle|\,\alpha\in I(0)\right\}.\tag{37}$$

设$\lambda_k\to0,\alpha_k\in I(0)$满足

$$\frac{g(\lambda_k,\alpha_k)-g(0,\alpha_k)}{\lambda_k}=\max_{\alpha}\left\{\frac{g(\lambda_k,\alpha)-g(0,\alpha)}{\lambda_k}\middle|\alpha\in I(0)\right\}. \tag{38}$$

不破坏一般性，可以认为 $\alpha_k\to\alpha_0\in I(0)$。

由第二章定理3.10，有

$$\frac{g(0,\alpha_k)-g(-\delta,\alpha_k)}{\delta}\leqslant\frac{g(\lambda_k,\alpha_k)-g(0,\alpha_k)}{\lambda_k}\leqslant\frac{g(\delta,\alpha_k)-g(0,\alpha_k)}{\delta}.$$

故由 $g(\lambda,\alpha)$的连续性及 I 的紧致性，差商

$$[g(\lambda_k,\alpha_k)/g(0,\alpha_k)]/\lambda_k$$

有界. 可以设 $\lambda_k\to 0$时它趋于一个数值 μ.

设 $\lambda>0$固定，则 k 充分大时，$\lambda>\lambda_k$. 由第二章定理 3.10 有

$$\frac{g(\lambda,\alpha_k)-g(0,\alpha_k)}{\lambda}\geqslant\frac{g(\lambda_k,\alpha_k)-g(0,\alpha_k)}{\lambda_k}.$$

在$k\to\infty$时取极限，得

$$\frac{g(\lambda,\alpha_0)-g(0,\alpha_0)}{\lambda}\geqslant\mu. \tag{39}$$

由此，令 $\lambda\downarrow 0$ 时得 $g'(0,\alpha_0)\geqslant\mu$. 利用(37)，(38)式得

$$g'(0)\leqslant\mu\leqslant g'(0,\alpha_0),\alpha_0\in I(0).$$

再根据(35)式，则

$$g'(0,\alpha_0)\geqslant g'(0)\geqslant\sup_{\alpha}\{g'(0,\alpha)\mid\alpha\in I(0)\},\alpha_0\in I(0).$$

即(33)式成立. ▌

定理 4.8 设 I 是紧致集. $f(x,\alpha)$在 $\alpha\in I$ 时是 x 的凸函数. 对于 x_0 的一个邻域及 $\alpha\in I$，$f(x,\alpha)$关于 x 和 α 连续，则有

$$\partial f(x_0)=\mathrm{cl}(\mathrm{co}\bigcup_{\alpha\in I(x_0)}\partial f(x_0,\alpha)). \tag{40}$$

证明 根据支撑函数与闭凸集的一一对应，只要证明(40)式两边集合的支撑函数相等即可.

由定理1.7知

$$f'(x_0:p,\alpha)=\max\{<p,x^*>|x^*\in\partial f(x_0,\alpha)\}. \qquad (41)$$

利用定理 4.7 的(32)式及(41)式，有

$$\begin{aligned} f'(x_0:p) &= \max_{\alpha\in I(x_0)} \max_{x^*\in\partial f(x_0,\alpha)} \langle p,x^*\rangle \\ &= \max_{x^*}\{\langle p,x^*\rangle \mid x^*\in\bigcup_{\alpha\in I(x_0)}\partial f(x_0,\alpha)\} \\ &= \sup_{x^*}\{\langle p,x^*\rangle \mid x^*\in \mathrm{cl}(\mathrm{co}\bigcup_{\alpha\in I(x_0)}\partial f(x_0,\alpha)\}, \end{aligned} \qquad (42)$$

(42)式的最后一步等式是因为线性函数在一个集合的上确界和在这个集合的凸包的上确界相等，线性函数在一个集 合 的上确界和在这个集合的闭包的上确界相等。

再由定理1.7，得

$$f'(x_0:p)=\sup_{x^*}\{\langle p,x^*\rangle \mid x^*\in\partial f(x_0)\}. \qquad (43)$$

由(42)，(43)式知(40)式两边集合的支撑函数 相 等，故(40) 式成立. ▌

习　　题

1．设$f(x)$是 R^n 上的凸函数，$f(x)$、$f(y)$有限．证明

$$f'(y:y-x)\geqslant f'(x:y-x).$$

2．设$C\subset R^n$是非空闭凸集．证明

$$\partial\delta^*(x^*|C)=\{x\mid\langle x,x^*\rangle=\max_{x\in C}\langle x,x^*\rangle\}.$$

3．设$K\subset R^n$是非空闭凸锥．证明 $x^*\in\partial\delta(x|K)$的充分必要条件是$x\in\partial\delta(x^*|\tilde{K})$。这个条件也等价于

$$x\in K,x^*\in\tilde{K}\ 时,\langle x,x^*\rangle=0.$$

4．设$f(x)$是 R^n 上的正常凸函数，$x\in\mathrm{int}(\mathrm{dom}f)$，$f(x)>\inf f(x)$．证明 x^* 是$C=\{z|f(z)\leqslant f(x)\}$在 x 的法线向量的充分必要条件是存在一个 $\lambda\geqslant0$，使$x^*\in\lambda\partial f(x)$。

5．在 4 题的条件下证明，$\partial f(x)$是不包含原点的非空有界闭凸集。

6. 设$f(x)$是 R^n 上的正常凸函数，$\lambda>0$. 证明

$$\partial(\lambda f)(x)=\lambda\partial f(x).$$

7. 证明定理1.4.

8. 证明推论1.5.1.

9. 证明定理2.1中的(3).

10. 设$x\in R$,令

$$f(x)=\begin{cases}|x|-2(1-x)^{1/2} & -3\leqslant x\leqslant 1,\\ +\infty & \text{其他}.\end{cases}$$

求 $\partial f(x)$.

11. 证明推论3.2.1.

12. 设$f(x)$是开区间 $I\subset R$ 上可微的实值函数. 证明$f(x)$ 是凸函数的充分必要条件是$f'(x)$在 I 中不减少.

13. 设凸函数$f(x)$在 x 有唯一次梯度 x^*,令

$$g(y)=f(x+y)-f(x)-\langle x^*,y\rangle.$$

证明凸函数 g 在原点有唯一次梯度 θ.

14. 证明推论4.2.1.

15. 证明推论4.4.1.

16. 设$f(x)$是 R^n 上的凸函数，$C=\{x\,|\,f(x)\leqslant 0\}$,且存在一点 $x_1\in C$, $f(x_1)<0$. 如果$f(x_0)=0$,证明

$$[\text{cone}(C-x_0)]^*=-\text{cone}\,\partial f(x_0).$$

17. 设 $C\subset R^n$ 是非空凸集，$f(x)$是 R^n 上的实值函数，且$\text{dom}\,f=C$. 如果对每一个 $\bar{x}\in\text{int}\,C$,存在 $y\in\partial f(\bar{x})$满足：

$$f(x)\geqslant f(\bar{x})+\langle y,x-\bar{x}\rangle,\ \forall x\in C.$$

证明$f(x)$是凸函数.

第四章　极值问题的最优性条件

极值问题是最优化理论的核心，这一章将在凸性和一般的情形下讨论极值问题的最优性条件.

§1　线性规划

求线性函数在由线性(仿射)不等式组定义的凸集上的极值问题称为线性规划(LP)，即

$$\inf\langle x, x_0^*\rangle, \qquad x\in R^n, \tag{1}$$

(LP)

$$\text{s.t}\langle x, x_i^*\rangle - a_i \leqslant 0, i=1,\cdots,m. \tag{2}$$

或

$$\inf_x\{\langle x, x_0^*\rangle \mid \langle x, x_i^*\rangle - a_i \leqslant 0, i=1,\cdots,m\}. \tag{3}$$

如果至少存在一点 x 满足(2)式，则称不等式组(2)相容. 设

$$D=\{x\mid\langle x, x_i^*\rangle - a_i \leqslant 0, i=1,\cdots,m\},$$

称 D 是可行区域. $\langle x, x_0^*\rangle$ 称为目标函数.

1. 极值存在的必要条件

定理 1.1　在(LP)中，目标函数的下确界达到且有限，或者等于 $-\infty$.

证明　根据第一章定理 9.18 中的(47)式，不等式组(2) 的解可以写成

$$x=\sum_{j\in I^+}\nu_j y_j+\sum_{j\in I^\circ}\lambda_j x_j, \tag{4}$$

$$\sum_{j\in I^+}\nu_j=1, \nu_j\geqslant 0, \lambda_j\geqslant 0, \tag{5}$$

其中 $I^\circ\cup I^+=\{1,\cdots,m\}$. 故对于 $\forall x\in D$，有

$$\langle x, x_0^*\rangle=\sum_{j\in I^+}\nu_j\langle y_j, x_0^*\rangle+\sum_{j\in I^\circ}\lambda_j\langle x_j, x_0^*\rangle. \tag{6}$$

于是(LP)归结为(6)式右边式子在 ν_j，λ_j 满足约束(5) 式的极小化.

如果对于某个 $j\in I^\circ$，$\langle x_j,x_0^*\rangle<0$，令相应的 $\lambda_j\to -\infty$，则(6) 式的右边趋于$-\infty$.

如果对于 $\forall j\in I^\circ$，$\langle x_j,x_0^*\rangle\geqslant 0$. 因为求下确界，可以令 $\lambda_j=0$，$j\in I^\circ$. 故

$$\min_x\{\langle x, x_0^*\rangle \mid x\in D\}$$

$$=\min_{\nu_j}\left\{\sum_{j\in I^+}\nu_j\langle y_j, x_0^*\rangle \mid \sum\nu_j=1,\ \nu_j\geqslant 0\right\}$$

$$=\min_{j\in I^+}\langle y_j, x_0^*\rangle. \tag{7}$$

所以最小值将在满足

$$\langle y_{j_0}, x_0^*\rangle=\min_{j\in I^+}\langle y_j, x_0^*\rangle \tag{8}$$

的 y_{j_0} 处达到且有限. ▌

从第一章定理 9.18 及(4)，(5)式知，$D=M_D+K_D$，其中 M_D 是多面体，K_D是多面锥. 而 y_j 是 M_D 的极点.

定理 1.2 线性规划问题(LP)的解集可以表示成

$$x=\sum_{j\in I_0^+}\nu_j y_j+\sum_{j\in I_0^\circ}\lambda_j x_j, \tag{9}$$

$$\sum_{j\in I_0^+}\nu_j=1, \nu_j\geqslant 0,\ \lambda_j\geqslant 0, \tag{10}$$

其中

$$I_0^+=\{j\in I^+ \mid \langle y_j, x_0^*\rangle=\min_{k\in I^+}\langle y_k, x_0^*\rangle\},$$

$$I_0^\circ=\{j\in I^\circ \mid \langle x_j, x_0^*\rangle=0\}.$$

证明 容易验证，表示成(9)式的 x 一定满足

$$\langle x, x_0^*\rangle=\min_{j\in I^+}\langle y_j, x_0^*\rangle.$$

由(8)式,$\langle x, x_0^*\rangle$在这样的 x 达到下确界,故 x 是(LP)的解.

如果 x 不能表示成(9)式的形式,则由(6) 式知

$$\langle x, x_0^*\rangle > \min_{i\in I^*}\langle y_i, x_0^*\rangle, \tag{11}$$

即 x 不是(LP)的解. ▌

推论 1.2.1 设 D 有界,则 D 是多面体,目标函数的极小值在 D 的极点的子集上达到,且任何极小值点均可表示成这些极点的凸组合.

证明留给读者.

定理 1.3 设 x_0 是(LP)的极小值点,则存在 $\lambda_i \geqslant 0$, $i=1,\cdots,m$, 满足

$$x_0^* + \sum_{i=1}^{m}\lambda_i x_i^* = \theta, \tag{12}$$

$$\lambda_i(\langle x_0, x_i^*\rangle - \alpha_i) = 0, i=1,\cdots,m. \tag{13}$$

反之,如果(12),(13)式满足,则 $x_0\in D$ 是(LP)的极小值点.

证明 设 x_0 是极小值点. 研究凸锥:

$$K_D(x_0) = \{\bar{x} \mid x_0 + \lambda\bar{x}\in D, \lambda \text{ 充分小}, \lambda>0\}. \tag{14}$$

设 $I_0=\{i \mid \langle x_0, x_i^*\rangle - \alpha_i = 0, i=1,\cdots,m\}$. 如果 $\bar{x}\in K_D(x_0)$, 则在$\lambda>0$ 充分小时,下面的不等式成立:对 $i\in I_0$, 有

$$\langle x_0+\lambda\bar{x}, x_i^*\rangle - \alpha_i = \langle x_0, x_i^*\rangle - \alpha_i + \lambda\langle\bar{x}, x_i^*\rangle \leqslant 0. \tag{15}$$

容易看出, (15)式成立的等价条件是:

$$\langle\bar{x}, x_i^*\rangle \leqslant 0, \quad i\in I_0.$$

于是

$$K_D(x_0) = \{\bar{x} \mid \langle\bar{x}, -x_i^*\rangle \geqslant 0, i\in I_0\}. \tag{16}$$

因为 x_0 是极小值点,而 $\bar{x}\in K_D(x_0)$时, 只要 λ 充分小, 有 $x_0 + \lambda\bar{x}\in D$,故

$$\langle x_0+\lambda\bar{x}, x_0^*\rangle \geqslant \langle x_0, x_0^*\rangle,$$

即

$$\langle x, x_0^*\rangle \geqslant 0, x \in K_D(x_0).$$

换言之，$x_0^* \in K_D^*(x_0)$. 根据第一章定理 9.14 及(16)，x_0^*可以表示成

$$x_0^* = -\sum_{i \in I_0} \lambda_i x_i^*,\ \lambda_i \geqslant 0, i \in I_0$$

如果在 $i \notin I_0$ 时，设 $\lambda_i = 0$. 于是(12)，(13)两式成立.

反之，设 $x_0 \in D$ 满足(12)，(13)式，则对 $\forall x \in D$，有

$$\begin{aligned}\langle x, x_0^*\rangle &\geqslant \langle x, x_0^*\rangle + \sum_{i=1}^m \lambda_i(\langle x, x_i^*\rangle - \alpha_i) \\ &= \langle x, x_0^* + \sum_{i=1}^m \lambda_i x_i^*\rangle - \sum_{i=1}^m \lambda_i \alpha_i \\ &= -\sum_{i=1}^m \lambda_i \alpha_i.\end{aligned} \tag{17}$$

同时，由(13)式，有

$$\begin{aligned}\langle x_0, x_0^*\rangle &= \langle x_0, x_0^*\rangle + \sum_{i=1}^m \lambda_i(\langle x_0, x_i^*\rangle - \alpha_i) \\ &= -\sum_{i=1}^m \lambda_i \alpha_i.\end{aligned} \tag{18}$$

从(17)，(18)式知，对 $\forall x \in D$，$\langle x, x_0^*\rangle \geqslant \langle x_0, x_0^*\rangle$，所以 x_0 是极小值点. ▌

2. 对偶问题

设 $y^* = (\lambda_1, \cdots, \lambda_m)$，$\varphi(y^*) = -\sum_{i=1}^m \lambda_i \alpha_i$，

$$D_* = \left\{y^* \middle| x_0^* + \sum_{i=1}^m \lambda_i x_i^* = \theta, \lambda_i \geqslant 0, i = 1, \cdots, m\right\}.$$

定义 1.1 求 $\varphi(y^*)$ 在 D_* 上的极大值问题称为(LP)的对偶问题. 用(DLP)表示.

定理 1.4 如果 $x \in D, y^* \in D_*$, 则 $\langle x, x_0^* \rangle \geq \varphi(y^*)$.

证明留给读者.

定理 1.5 如果 (LP) 的目标函数的下确界等于 $-\infty$, 则(DLP)的约束不相容,即 $D_* = \varnothing$.

证明留给读者.

定理 1.6 如果(LP)有解 x_0,则(DLP)有解 y_0^*, 且 $\langle x_0, x_0^* \rangle = \varphi(y_0^*)$.

证明 如果取定理 1.3 中的 λ_i 作为 y_0^* 的各个分量, $i=1,\cdots,m$. 则由定理 1.4 和(18)式, 得

$$\varphi(y_0^*) = \langle x_0, x_0^* \rangle \geq \varphi(y^*), \ y^* \in D_*. \tag{19}$$

则 y_0^* 是(DLP)的解,且 $\langle x_0, x_0^* \rangle - \varphi(y_0^*)$. ▌

定理 1.7 如果 $x \in D, y^* \in D_*$. 则 x 是线性规划问题的解, y^* 是其对偶问题的解的等价条件是

$$\lambda_i(\langle x, x_i^* \rangle - \alpha_i) = 0, i = 1, \cdots, m, \tag{20}$$

其中 $y^* = (\lambda_1, \cdots, \lambda_m)$.

证明留给读者.

3. 线性不等式组

先讨论齐次线性不等式组:

$$\langle x, x_i^* \rangle \leq 0, \ i = 1, \cdots, m. \tag{21}$$

设 K_i 是(21)式中第 i 个不等式定义的闭半空间. 显然, int K_i 由满足 $\langle x, x_i^* \rangle < 0$ 的 x 组成, $\mathrm{Lin}\, K_i = R^n$.

定理 1.8 不等式组:

$$\langle x, x_i^* \rangle < 0, i = 1, \cdots, m, \tag{22}$$

有解的等价条件是不存在不全为 0 的 $\lambda_i \geq 0, i = 1, \cdots, m$, 满足

$$\sum_{i=1}^{m}\lambda_i x_i^*=\theta. \tag{23}$$

证明 利用第一章定理 8.14 研究前面所定义的 K_i的不可分离性.

因为 $\mathrm{Lin}\,K_i=R^n$,故那里的条件 2)自然满足,所以这些凸锥不可分离的等价条件是 $\mathrm{ri}K_1\cap\cdots\cap\mathrm{ri}K_m\neq\varnothing$,而这等价于(22)有解.

因为 $K_i^*=\{-\lambda_i x_i^*|\lambda_i\geqslant 0\}$,根据第一章凸锥不可分离的定义 8.4,知(22)有解的等价条件是(23)式不能成立. ▌

定理 1.9 不等式组:

$$\langle x,x_i^*\rangle\leqslant 0,\ i=1,\cdots,m, \tag{24}$$
$$\langle x,x^*\rangle<0,$$

相容的等价条件是不存在 $\lambda_i\geqslant 0,i=1,\cdots,m$,满足

$$x^*+\sum_{i=1}^{m}\lambda_i x_i^*=\theta. \tag{25}$$

证明 (24)式不相容的等价条件是 x 是不等式组(21)的解时,$\langle x,x^*\rangle\geqslant 0$,即 $x^*\in K^*$,其中K是由(21)式定义的凸锥. 但由第一章定理 9.14,有

$$x^*=-\sum_{i=1}^{m}\lambda_i x_i^*,\ \lambda_i\geqslant 0,\ i=1,\cdots,m. \tag{26}$$

即(25)式成立. 故(24)式相容的等价条件是(25)式不成立. ▌

推论 1.9.1 不等式组(21)具有唯一解 $x=\theta$ 的充分必要条件是对于 $\forall x^*\in R^n$,存在 $\lambda_i\geqslant 0$,使(25)式成立.

证明留给读者.

下面讨论非齐次不等式组:

$$\langle x,x_i^*\rangle-\alpha_i\leqslant 0,\ i=1,\cdots,m. \tag{27}$$

容易看出,(27)的解的存在性归结为不等式组

$$\langle x, x_i^* \rangle - \alpha_i x^0 \leqslant 0, i=1, \cdots, m, \tag{28}$$
$$-x^0 < 0.$$

的解的存在性的研究，其中$(x, x^0) \in R^{n+1}$是未知向量.

定理 1.10 不等式组：

$$\langle x, x_i^* \rangle - \alpha_i < 0, i=1, \cdots, m.$$

有解的等价条件是不存在不全为 0 的$\lambda_i \geqslant 0$,满足

$$\sum_{i=1}^{m} \lambda_i x_i^* = \theta, \sum_{i=1}^{m} \lambda_i \alpha_i \leqslant 0. \tag{29}$$

证明留给读者.

定理 1.11 不等式组(27) 相容的等价条件是不存在$\lambda_i \geqslant 0$, $i=1, \cdots, m$,满足

$$\sum_{i=1}^{m} \lambda_i x_i^* = \theta, \sum_{i=1}^{m} \lambda_i \alpha_i = -1. \tag{30}$$

证明留给读者.

定理 1.12 设K是(21)式定义的凸锥. 则(27) 的解有界的等价条件是$K = \{\theta\}$.

证明 设$x_0 \in K, x_0 \neq \theta$, x满足(27)式,则对$\forall \rho \geqslant 0, x + \rho x_0$也满足(27)式，故(27)的解无界.

反之,设x_j是(27)的解. $j=1,2,\cdots, |x_j| \xrightarrow[j\to\infty]{} +\infty$, 则从$\{x_j\}$中选取收敛序列$y_j = |x_j|^{-1} x_j, y_j \to y_0, |y_0| = 1$,得

$$\langle y_j, x_i^* \rangle - \frac{\alpha_i}{|x_j|} \leqslant 0, i=1, \cdots, m.$$

令$j \to \infty$, 得

$$\langle y_0, x_i^* \rangle \leqslant 0, i=1, \cdots, m.$$

故 $y_0 \in K$,且$y_0 \neq \theta$. ∎

推论 1.12.1 (27) 的解集有界的充分必要条件是对于$\forall x^* \in R^n$,存在满足(25)式的$\lambda_i \geqslant 0, i=1, \cdots, m$.

证明留给读者.

定理 1.13 设(27)式不相容,则存在它的一个子不等式组:

$$\langle x, x_i^*\rangle - \alpha_i \leqslant 0, i \in I, I \subset \{1, 2, \cdots, m\},$$

也不相容. 其中指标集 I 的元素个数不超过 $n+1$.

证明 设(27)式不相容,则由定理 1.11,存在 $\lambda_i \geqslant 0$,满足

$$\sum_{i=1}^{m} \lambda_i x_i^* = \theta, \sum_{i=1}^{m} \lambda_i \alpha_i = -1. \tag{31}$$

设

$$J = \{i \mid \lambda_i > 0, i = 1, \cdots, m\}. \tag{32}$$

如果 J 的元素的个数大于 $n+1$,则向量$(x_i^*, \alpha_i) \in R^{n+1}$ 线性相关,故存在不全为 0 的数 ν_i,满足

$$\sum_{i \in J} \nu_i x_i^* = \theta, \sum_{i \in J} \nu_i \alpha_i = 0. \tag{33}$$

设

$$\varepsilon_0 = \min_{i \in J} \left\{ \frac{\lambda_i}{\nu_i} \middle| \nu_i > 0 \right\},$$

则 $\overline{\lambda}_i = \lambda_i - \varepsilon_0 \nu_i$ 非负,且由(31)(33)式得

$$\sum_{i \in J} \overline{\lambda}_i x_i^* = \theta, \sum_{i \in J} \overline{\lambda}_i \alpha_i = -1. \tag{34}$$

由 ε_0 的选择,至少有一个$\overline{\lambda}_i = 0, i \in J$. 故(34)式中至少有一项为 0. 于是从中除去为 0 的项并重复上述过程,则在有限次内(34)式的非零项将不多于 $n+1$. 取非零的 λ_i 的指标 i 作成指标集I,根据定理 1.11 知,I 所对应的子不等式组不相容. ∎

定理 1.13 实际说明,(27)式相容的等价条件是它的任意 $n+1$ 个不等式构成的子不等式组相容.

最后,证明在最优性条件中广泛应用的两个定理.

定理 1.14 (Farkas定理) 设 A 是 $m \times n$ 矩阵,c 是 n 维向量. 则下面两个结论恰有一个成立:

1) 存在 $x\in R^n$,满足 $Ax\leqslant\theta,\langle c,x\rangle>0$. (35)

2) 存在 $y\in R^m$,满足 $y\geqslant\theta,A^*y=c$. (36)

证明 设 2) 成立,即存在 $y\in R^m$,使 $y\geqslant\theta$, $A^*y=c$. 如果 $x\in R^n$ 使 $Ax\leqslant\theta$,则 $\langle c,x\rangle=\langle y,Ax\rangle\leqslant 0$,故 1)不成立.

设 2)不成立. 令 $S_1=\{c\},S_2=\{x\,|\,x=A^*y,y\geqslant\theta\}$, 则 S_2 是闭凸集,$S_1\cap S_2=\varnothing$. 由第一章定理 5.8,S_1、S_2可以强分离. 故由第一章定理 5.6,存在向量 $p\in R^n$和实数 α, 满足:

$$\langle p,c\rangle>\alpha,\ \alpha\geqslant\sup_{x\in S_2}\langle p,x\rangle.$$

因为 $\theta\in S_2$,故 $\alpha\geqslant 0$,而 $\alpha\geqslant\langle p,A^*y\rangle=\langle Ap,y\rangle$, $y\geqslant\theta$. 但 y 的分量可以任意大,所以 $Ap\leqslant\theta$. 最后有

$$\langle c,p\rangle>0,Ap\leqslant\theta. \tag{37}$$

即 1)成立. ▍

Farkas 定理也可叙述为: 设 A 是 $m\times n$ 矩阵, $c\in R^n$. 则系统 $Ax\leqslant\theta,\langle c,x\rangle>0$ 有解的等价条件是系统 $A^*y=c,y\geqslant\theta$ 无解.

定理 1.14 的几何解释是:设 $a_1,\cdots,a_m$ 是A^*的 m 个 n 维列向量. 如果c在由$a_1,\cdots,a_m$生成的凸锥之中,则闭凸锥$\{x\,|\,Ax\leqslant\theta\}$与开半空间$\{x\,|\,\langle c,x\rangle>0\}$的交是空集. 反之亦然. 如图 1.

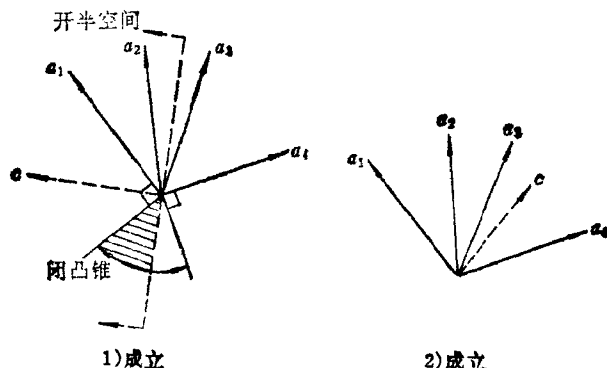


图 1 Farkas 定理的几何解释

推论 1.14.1 设A是$m\times n$矩阵, $c\in R^n$, 则下面两个结论恰

有一个成立：

1)存在$x\in R^n$,满足$Ax\leqslant\theta,\langle c、x\rangle>0,x\geqslant\theta$. (38)

2)存在$y\in R^m$,满足$y\geqslant\theta,A^*y\geqslant c$. (39)

证明 设$B^*=(A^*,-I)$,其中I为单位矩阵. 如果1)成立，则存在$x\in R^n$,满足

$$Bx=\begin{pmatrix}A\\-I\end{pmatrix}x=\begin{pmatrix}Ax\\-x\end{pmatrix}\leqslant\theta,\langle c、x\rangle>0. \tag{40}$$

反之,如果$x\in R^n$满足$Bx\leqslant\theta,\langle c、x\rangle>0$,则必有

$$Ax\leqslant\theta,x\geqslant\theta,\langle c、x\rangle>0.$$

所以1) 与定理1.14的1) 等价.

设$w=(y,z),y\in R^m,z\in R^n$. 如果2)成立,令$A^*y-c=z$. 则$z\geqslant\theta$. 且

$$B^*w=(A^*,-I)w=A^*y-z=c,w\geqslant\theta. \tag{41}$$

反之,如果$w\in R^{n+m}$满足$B^*w=c,w\geqslant\theta$,则y为2) 的解. 所以2) 与定理1.14的2)等价,故推论的结论成立. ▌

推论 1.14.2 设A是$m\times n$矩阵,B是$l\times n$矩阵,$c\in R^n$. 则下面两个结论恰有一个成立：

1)存在$x\in R^n$,满足$Ax\leqslant\theta,Bx=\theta,\langle c,x\rangle>0$. (42)

2)存在$y\in R^m,z\in R^l$,满足$A^*y+B^*z=c,y\geqslant\theta$. (43)

证明 用$(A^*,B^*,-B^*)$代替定理1.14中的A^*,与推论1.14.1的证明方法类似. ▌

定理 1.15 (Gordan 定理) 设A是$m\times n$矩阵,则下面两个结论恰有一个成立：

1)存在$x\in R^n$,满足$Ax<\theta$. (44)

2)存在非零的$y\in R^m$,满足$A^*y=\theta,y\geqslant\theta$. (45)

证明 这是定理1.8的变形. ▌

§ 2 约束极值问题

本节研究形式为

$$\min f(x) \tag{1}$$
$$\text{s.t} \quad x \in S.$$

的极值问题，其中 $f(x)$ 是 R^n 上的实值函数，$S \subset R^n$. 在很多情形下，$f(x)$ 是非线性函数，S 由一组非线性等式和不等式组确定，所以也称为非线性规划问题（NLP）. 显然，线性规划问题是（NLP）的特殊情形.

定义 2.1 设 $f(x)$ 是 $S \subset R^n$ 上的实值函数. 如果存在一个 $\delta > 0$，对于 $\forall x \in \{x \mid |x - x_0| < \delta\} \cap S$，不等式

$$f(x_0) \leq f(x) \tag{2}$$

成立，则称 x_0 是 $f(x)$ 的局部极小值点，$f(x_0)$ 是 $f(x)$ 的局部极小值. 如果对于 $\forall x \in S$，均有

$$f(x_0) \leq f(x), \tag{3}$$

则称 x_0 是 $f(x)$ 的全局极小值点，$f(x_0)$ 是 $f(x)$ 的全局极小值. 如果(2)，(3) 中的不等号是严格的，则称 x_0 是严格局部(全局) 极小值点.

1. 无约束的情形

这是 $S = R^n$ 的情形. 设 $\text{dom} f = D$.

定理 2.1 设 $f(x)$ 是 R^n 上的实值函数，在 $x^* \in D$ 可微. 如果存在 $y \in R^n$，满足 $\langle \nabla f(x^*), y \rangle < 0$，则存在 $\delta > 0$，对于 $\forall \lambda \in (0, \delta)$，有 $f(x^* + \lambda y) < f(x^*)$.

证明 根据 $f(x)$ 在 x^* 的可微性，有

$$f(x^* + \lambda y) = f(x^*) + \langle \nabla f(x^*), \lambda y \rangle + o(|\lambda y|),$$

即

$$\frac{f(x^* + \lambda y) - f(x^*)}{\lambda} = \langle \nabla f(x^*), y \rangle + \frac{1}{\lambda} o(|\lambda y|). \tag{4}$$

因为$\langle \nabla f(x^*), y \rangle < 0$，所以存在$\delta > 0$，当$\lambda \in (0, \delta)$时，

$$\langle \nabla f(x^*), y \rangle + \frac{1}{\lambda} o(|\lambda z|) < 0.$$

由(4) 知，$f(x^* + \lambda y) < f(x^*)$. ▌

上面定理中的y称为$f(x)$ 在x^*的一个下降方向.

推论 2.1.1 设$f(x)$ 是R^n上的实值函数，在x^*可微. 如果x^*是$f(x)$ 的局部极小值点，则$\nabla f(x^*) = \theta$.

证明 设$\nabla f(x^*) \neq \theta$，$y = -\nabla f(x^*)$，则$\langle \nabla f(x^*), y \rangle = -|\nabla f(x^*)|^2 < 0$. 由定理 2.1，存在$\delta > 0$，在$\lambda \in (0, \delta)$时，$f(x^* + \lambda y) < f(x^*)$. 这与$x^*$是局部极小值点矛盾，故$\nabla f(x^*) = \theta$. ▌

上面所述的 $f(x)$ 达到局部极小的必要条件使用了梯度向量，称为一阶条件. 下面的定理利用Hesse矩阵来叙述，也称为二阶条件.

定理 2.2 设$f(x)$ 是 R^n上的实值函数，在 x^*二阶连续可微. 如果$f(x)$ 在x^*取得局部极小值，则

1) $\nabla f(x^*) = \theta$, (5)

2) $f''(x^*)$ 半正定，即对$\forall z \in R^n$，有

$$\langle z, f''(x^*) z \rangle \geq 0. \tag{6}$$

证明 设$f(x)$ 在x^*取得局部极小值. 由定义 2.1，对于$\forall x \in \{x \mid |x - x^*| < \delta\} \cap D$，均有$f(x) \geq f(x^*)$. 对这些$x$，令$x = x^* + \lambda y$，$\lambda \in R$，$y \in B$，其中$B$是单位球. 故当$|\lambda|$充分小时，$f(x^* + \lambda y) \geq f(x^*)$.

设y固定，令$F(\lambda) = f(x^* + \lambda y)$. 则在$|\lambda| < \delta$时

$$F(\lambda) \geq F(0). \tag{7}$$

由中值定理，$F(\lambda) = F(0) + F'(\rho\lambda)\lambda$，$0 < \rho < 1$.

如果$F'(0) > 0$，则由 $f(x)$ 连续可微的假设，对于满足 $0 < \rho < 1$ 的ρ及$|\lambda| < \delta$，有$f'(\rho\lambda) > 0$. 故存在一个$\lambda < 0$，$|\lambda| < \delta$，使$F(0) > F(\lambda)$，这与(7)式矛盾. 相似可以证明$F'(0) < 0$ 也是不可

能的．所以$F'(0)=\langle y,\nabla f(x^*)\rangle=0$．由$y$在$B$中的任意性知$\nabla f(x^*)=\theta$，(5) 式成立．

由Taylor 定理，有

$$F(\lambda)=F(0)+F'(0)\lambda+\frac{1}{2}F''(\rho\lambda)\lambda^2,\ 0<\rho<1. \tag{8}$$

如果$F''(0)<0$，则由二阶连续可微的条件，存在$\varepsilon'>0$，当$0<\rho<1$及$|\lambda|<\varepsilon'$时，$F''(\rho\lambda)<0$．而$F'(0)=0$，故$F(\lambda)<F(0)$，与(7) 式矛盾．所以，有

$$F''(0)=\langle y,f''(x^*)y\rangle\geqslant 0.$$

因为$\forall z\in R^n$均可表为$z=ky,k\in R,y\in B,B$是单位球，故对一切$z\in R^n$，$\langle z,f''(x^*)z\rangle\geqslant 0$，故(6)式成立．▌

定理 2.3 设$f(x)$是R^n上的实值函数，在x^*二阶连续可微．如果$\nabla f(x^*)=\theta$，且对于$\forall x\in\{x\,|\,|x^*-x|<\delta\}$及$\forall z\in R^n$，有

$$\langle z,f''(x)z\rangle\geqslant 0, \tag{9}$$

则$f(x)$在x^*达到局部极小值．

证明 设x^*不是局部极小值点，则存在一个$w\in\{x\,|\,|x^*-x|<\delta\}$，使$f(x^*)>f(w)$．

令$w=x^*+\lambda y$，其中$y\in B,B$是单位球，$\lambda>0$．则由 Taylor 定理，有

$$f(w)=f(x^*)+\lambda\langle y,\nabla f(x^*)\rangle+\frac{1}{2}\lambda^2\langle y,f''(x^*+\lambda\rho y)y\rangle,$$

$0<\rho<1$．由$\nabla f(x^*)=\theta$，故有

$$\langle y,f''(x^*+\lambda\rho y)y\rangle<0. \tag{10}$$

因为$x^*+\lambda\rho y\in\{x\,|\,|x^*-x|<\delta\}$，这与(9) 式矛盾．所以$x^*$是$f(x)$的局部极小值点．▌

用与定理 2.3 相似的办法可以证明下面的结果．

定理 2.4 设$f(x)$是R^n上的实值函数，$x^*\in\text{int}D$，$f(x)$在

x^*二阶连续可微. 如果$\nabla f(x^*)=\theta$,且对于x^*的一个邻域中的$\forall x\neq x^*$和任意非零的$z\in R^n$,有

$$\langle z, f''(x)z\rangle>0, \tag{11}$$

则x^*是$f(x)$ 的严格局部极小值点.

例 1 设$f(x)=x^{2p}$,$x\in R$.p是正整数. $f'(x)=2px^{2p-1}$, $f'(0)=0$. 故在$x=0$满足一阶必要条件.

$$f''(x)=2p(2p-1)x^{2p-2}.$$

1) $p=1$时,$f''(0)=2$. 故$x=0$是严格极小值点

2) $p>1$时,$f''(0)=0$. 但容易检验,在$x=0$处定理2.3、定理2.4中的条件(9)、(11)都满足,故$x=0$是局部严格极小值点. 实际上它是全局严格极小值点,如图2.

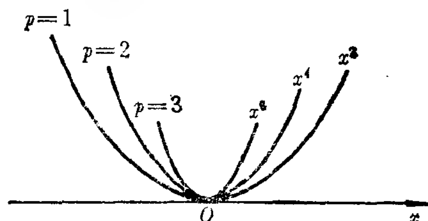


图 2 在原点邻域中的函数$f(x)=x^{2p}$

2. 不等式和等式约束的情形

研究

$$\begin{aligned} &\min \quad f(x) && (12)\\ (\text{NLP})\quad &\text{s.t}\quad g_i(x)\leqslant 0,\quad i=1,\cdots,m, && (13)\\ &\qquad\ \ h_j(x)=0,\quad j=1,\cdots,p. && (14)\end{aligned}$$

其中$f(x)$,$g_i(x)$,$h_j(x)$是在R^n的开子集S中可微的实值函数. 如果不存在(14)式形式的约束,称为具有不等式约束的非线性规划问题(NLP_1),而把存在(13)、(14)两种形式约束的问题称为具有不等式和等式约束的非线性规划问题(NLP_2). 用X表示满足约束的集合,称为可行集,$x\in X$称为可行点.

定义 2.2 如果在(NLP)的可行点 x，(13)式的第 i 个约束函数 $g_i(x)=0$，则称该不等式约束是关于可行点 x 的起作用约束，或紧约束．否则，称为不起作用约束，或松约束．

用$I(x)=\{i \mid g_i(x)=0\}$表示在x的紧约束的指标集．

定义 2.3 称集合$D(x)=\{y \mid y\neq\theta,\ x+\lambda y\in X,\ x\in X,\ \forall\lambda\in(0,\delta)\}$是可行集$X$在$x$的可行方向锥，$y\in D(x)$称为在$x$的一个可行方向．

定理 2.5 在(NLP)中，如果$y\in D(x)$，$k\in I(x)$，则

$$\langle y,\nabla g_k(x)\rangle\leqslant 0. \qquad (15)$$

证明 由$g_k(x)$的可微性，知

$$g_k(x+\lambda y)=g_k(x)+\lambda\langle y,\nabla g_k(x)\rangle+o(|\lambda y|),\ \lambda>0.$$

如果$\langle y,\nabla g_k(x)\rangle>0$，则当$\lambda$充分小时，$\langle y,\nabla g_k(x)\rangle+o(|\lambda y|)>0$．但$g_k(x)=0$，故对充分小的$\lambda$，$g_k(x+\lambda y)>0$．这与$y$是可行方向的条件矛盾．故(15)式成立．▌

相似可以证明，如果$y\in D(x)$，则$\langle y,\nabla h_j(x)\rangle=0$，$j=1,\cdots,p$．故有下面的定义．

定义2.4 在(NLP)中，设$x\in X$．称

$$E(x)=\{y \mid \langle y,\nabla g_i(x)\rangle\leqslant 0,\ i\in I(x),\ \langle y,\nabla h_j(x)\rangle=0,\ j=1,\cdots,p\}$$

是X在x的线性化锥．

用$F(x)=\{y \mid \langle y,\nabla f(x)\rangle<0\}$表示$f(x)$在$x$的下降方向所组成的集合．

定理2.6 设$x_0\in X$，则$E(x_0)\cap F(x_0)=\varnothing$的充分必要条件是存在$\lambda_i\geqslant 0$，$i=1,\cdots,m$，$\mu_j$，$j=1,\cdots,p$，满足

$$\nabla f(x_0)+\sum_{i=1}^{m}\lambda_i\nabla g_i(x_0)+\sum_{j=1}^{p}\mu_j\nabla h_j(x_0)=\theta, \qquad (16)$$

$$\lambda_i g_i(x_0)=0, i=1,\cdots,m. \tag{17}$$

证明 因为$E(x_0)\neq\varnothing$. 故$E(x_0)\cap F(x_0)=\varnothing$ 的充分必要条件是对于满足

$$\langle y,\nabla g_i(x_0)\rangle\leqslant 0, i\in I(x_0), \tag{18}$$

$$\langle y,\nabla h_j(x_0)\rangle=0, j=1,\cdots,p \tag{19}$$

的每一个y，一定有

$$\langle y,\nabla f(x_0)\rangle\geqslant 0. \tag{20}$$

(19)式可以写成

$$\langle y,\nabla h_j(x_0)\rangle\geqslant 0, \tag{21}$$

$$\langle y,-\nabla h_j(x_0)\rangle\geqslant 0. \tag{22}$$

(18)式可以写成

$$\langle y,-\nabla g_i(x_0)\rangle\geqslant 0, i\in I(x_0). \tag{23}$$

根据定理 1.14(Farkas 定理)的等价形式. 对于满足(21)-(23)式的y,(20)式成立的充分必要条件是存在$\lambda_i\geqslant 0, i\in I(x_0)$, $\mu_j^1\geqslant 0, \mu_j^2\geqslant 0, j=1,\cdots,p$,使

$$\nabla f(x_0)=-\sum_{\lambda_i\in I(x_0)}\lambda_i\nabla g_i(x_0)+\sum_{j=1}^{p}(\mu_j^1-\mu_j^2)\nabla h_j(x_0) \tag{24}$$

如果当$i\notin I(x_0)$时，设$\lambda_i=0$，$\mu_j=\mu_j^2-\mu_j^1$，则由(24)知$E(x_0)\cap F(x_0)=\varnothing$的充分必要条件是(16),(17)式成立. ▌

从定理 2.6 知道，如果在(NLP)的局部极小点x^*能保证$E(x^*)\cap F(x^*)=\varnothing$,(16),(17)式就可能成为最优性必要条件. 虽然对于大多数(NLP)这个结论是成立的，但仍有例外的情形.

例2 研究R^2中的(NLP)：

$$\min f(x)=-\xi_1$$

$$\text{s.t}\quad g_1(x)=-(1-\xi_1)^3+\xi_2\leqslant 0,$$

$$g_2(x)=-\xi_1\leqslant 0,$$

$$g_3(x) = -\xi_2 \leqslant 0, x = (\xi_1, \xi_2).$$

$x_0 = (1,0)$是可行点，容易验证

$$E(x_0) = \{(y_1, y_2) \mid y_2 = 0\},$$

$$F(x_0) = \{(y_1, y_2) \mid y_1 > 0\}.$$

故$E(x_0) \cap F(x_0) \neq \varnothing$，不存在$\lambda_1, \lambda_2, \lambda_3$满足(16)、(17)．但 x_0 是局部极小值点．见图 3．

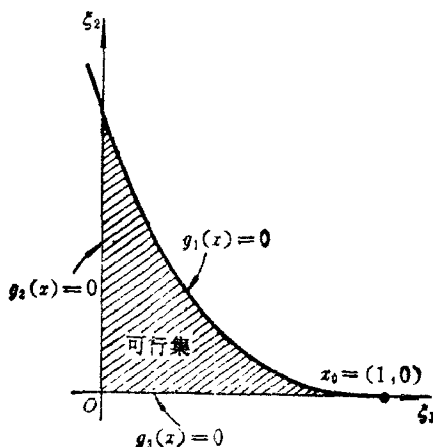


图 3　最优点没有相应乘子的例子

鉴于存在上述的例外情形，为了得到类似的一阶必要条件，只有考虑目标函数的梯度引入一个乘子λ_0，而使(16)变为

$$\lambda_0 \nabla f(x_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(x_0) = \theta. \tag{16$'$}$$

我们这样做自然是受到数学分析中条 件极值的 Lagrange 乘子法的启发．

仍然从几何角度开始．

定理2.7 设 x^*是（NLP）的局部极小值点. 则 $D(x^*)\cap F(x^*)=\varnothing$.

证明 设存在$y\in D(x^*)\cap F(x^*)$. 根据定理 2.1,存在$\delta_1>0$,使

$$f(x^*+\lambda y)<f(x^*),\forall\lambda\in(0,\delta_1). \tag{25}$$

由$D(x^*)$的定义,存在$\delta_2>0$,使

$$x^*+\lambda y\in X,\forall\lambda\in(0,\delta_2). \tag{26}$$

这与x^*是局部极小值点的条件矛盾,故$D(x^*)\cap F(x^*)=\varnothing$. ▌

定理 2.7 说明了这样一个事实:局部极小值点的一个必要条件是每个下降方向都不是可行方向. 见图 4.

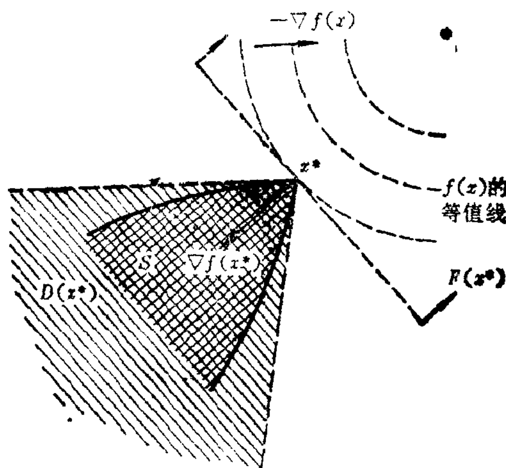


图 4 $D(x^*)\cap F(x^*)=\varnothing$的几何形象

为了叙述的简洁,下面仅讨论(NLP_1),即

$$\min f(x) \tag{27}$$

$$(NLP_1)\qquad \text{s.t}\quad g_i(x)\leqslant 0,i=1,\cdots,m \tag{28}$$

$$x\in S.$$

其中S是非空开集,这时可行集为

$$X=\{x \mid x\in S, g_i(x)\leqslant 0, i=1,\cdots,m\}.$$

设 $G(x)=\{y \mid \langle y,\nabla g_i(x)\rangle<0, i\in I(x)\}$.

定理2.8 设x^*是(NLP$_1$)的局部极小值点，则 $G(x^*)\cap F(x^*)=\varnothing$.

证明 设$y\in G(x^*)$. 因为S是开集，故存在$\delta_1>0$，使

$$x^*+\lambda y\in S, \lambda\in(0,\delta_1). \tag{29}$$

如果$i\notin I(x^*)$，$g_i(x^*)<0$. 则存在$\delta_2>0$，使

$$g_i(x^*+\lambda y)<0, \lambda\in(0,\delta_2), i\notin I(x^*). \tag{30}$$

最后由$y\in G(x^*)$，故对 $\forall i\in I(x^*)$，$\langle y, \nabla g_i(x^*)\rangle<0$. 由定理2.1，存在$\delta_3>0$，使

$$g_i(x^*+\lambda y)<g_i(x^*)=0. \ \lambda\in(0,\delta_3), i\in I(x^*). \tag{31}$$

取 $\delta=\min(\delta_1, \delta_2, \delta_3)$，故对$\lambda\in(0, \delta)$时，由(29)-(31)式知$x^*+\lambda y\in X$. 所以$y\in D(x^*)$，$G(x^*)\subset D(x^*)$. 但由定理 2.7，$x^*$是(NLP$_1$)的局部极小值点时，$D(x^*)\cup F(x^*)=\varnothing$. 所以$G(x^*)\cap F(x^*)=\varnothing$. ▮

例3 研究 R^2 中的(NLP$_1$)：

$$\begin{aligned}
&\min(\xi_1-3)^2+(\xi_2-2)^2\\
&\text{s.t} \quad \xi_1^2+\xi_2^2\leqslant 5,\\
&\qquad \xi_1+\xi_2\leqslant 3,\\
&\qquad -\xi_1\leqslant 0,\\
&\qquad -\xi_2\leqslant 0, \qquad x=(\xi_1,\xi_2).
\end{aligned}$$

在此，令 $g_1(x)=\xi_1^2+\xi_2^2-5$，$g_2(x)=\xi_1+\xi_2-3$，$g_3(x)=-\xi_1$，$g_4(x)=-\xi_2$，$S=R^2$.

考虑点 $\bar{x}=\left(\frac{9}{5}, \frac{6}{5}\right)$，这时$I(\bar{x})=\{2\}$，且

$$\nabla f(\bar{x})=\left(-\frac{12}{5},-\frac{8}{5}\right), \nabla g_2(\bar{x})=(1,1).$$

图 5 中的$G(\bar{x})$，$F(\bar{x})$已将顶点平移到$\bar{x}$，因为$G(\bar{x})\cap F(\bar{x})\neq\varnothing$，故 $\bar{x}=\left(\frac{9}{5},\ \frac{6}{5}\right)$不是上述问题的局部极小值点.

再考虑$\bar{\bar{x}}=(2,1)$，这时 $I(\bar{\bar{x}})=\{1,2\}$，且

$$\nabla f(\bar{\bar{x}})=(-2,-2),\nabla g_1(\bar{\bar{x}})=(4,2),\nabla g_2(\bar{\bar{x}})=(1,1).$$

图内表明 $G(\bar{\bar{x}})\cap F(\bar{\bar{x}})=\varnothing$. 但定理 2.8 仅给出必要条件，故不能保证$\bar{\bar{x}}$是局部极小值点.

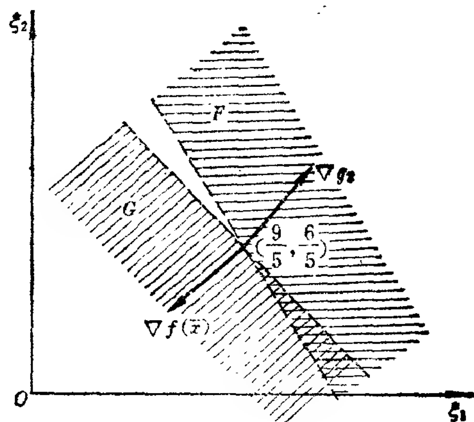


图 5　在非最优点$\bar{x}$，$G(\bar{x})\cap F(\bar{x})\neq\varnothing$.

其实，定理 2.8 的必要条件也可能被非局部极小值点平凡地满足. 下面就是这样的情形.

设x_0是一个可行点，$\nabla f(x_0)=\theta$. 显然，$F(x_0)=\varnothing$. 故$G(x_0)\cap F(x_0)=\varnothing$，从而满足$\nabla f(x)=\theta$的任何可行点都满足最优性必要条件. 同样，对某个$i\in I(x)$，使$\nabla g_i(x)=\theta$的任意点$x$也都满足必要条件. 下面是另一个例子.

考虑只有一个等式约束的情形：

$$\min f(x)$$
$$\text{s.t}\quad g(x)=0.$$

因为$g(x)=0$可以换为$g_1(x)=g(x)\leqslant 0$, $g_2(x)=-g(x)\leqslant 0$. 令x_0是一个可行点，则$g_1(x_0)=g_2(x_0)=0$，但$\nabla g_1(x_0)=-\nabla g_2(x_0)$，所以不存在$\langle y,\nabla g_1(x_0)\rangle<0$, $\langle y,\nabla g_2(x_0)\rangle<0$的向量$y$，$G(x_0)=\varnothing$，故$G(x_0)\cap F(x_0)=\varnothing$. 换言之 全部可行点都满足最优性必要条件.

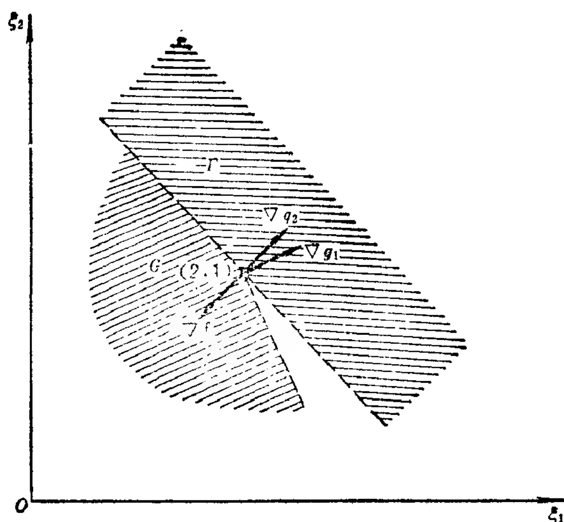


图 6 在最优点，$\bar{x}$ $G(\bar{x})\cap F(\bar{x})=\varnothing$.

现在将几何最优性必要条件$G(x^*)\cap F(x^*)=\varnothing$用目标函数和约束函数的梯度来表示. 下面的结果是 Fritz John 在 1948 年给出的.

定理2.9 (Fritz John条件) 如果x^*是(NLP_1)的局部极小值点，则存在不全为 0 的实数$\lambda_0,\lambda_1,\cdots,\lambda_m$，满足

$$\lambda_0\nabla f(x^*) \quad \sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)=\theta, \tag{32}$$

$$\lambda_i g_i(x^*)=0, i=1,\cdots,m \tag{33}$$

$$\lambda_0\geqslant 0,\lambda_i\geqslant 0, i=1,\cdots m.$$

证明 设x^*是局部极小值点，则由定理2.8，对于$\forall i\in I(x^*)$，不存在向量y满足

$$\langle y,\nabla f(x^*)\rangle<0,$$

$$\langle y,\nabla g_i(x^*)\rangle<0, i\in I(x^*).$$

设A是由$\nabla f(x^*)$，$\nabla g_i(x^*)$为其行的矩阵，故$Ay<\theta$不相容．根据定理1.15，存在非零向量$d\geqslant\theta$，满足$A^*d=\theta$．用λ_0，λ_i，$i\in I(x^*)$分别表示d的分量，则有

$$\lambda_0\nabla f(x^*)+\sum_{i\in I(x^*)}\lambda_i\nabla g_i(x^*)=\theta, \tag{34}$$

$$\lambda_i g_i(x^*)=0, i\in I(x^*). \tag{35}$$

在$i\notin I(x^*)$时，令$\lambda_i=0$，则(34)，(35)变成

$$\lambda_0\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)=\theta, \tag{36}$$

$$\lambda_i g_i(x^*)=0, i=1,\cdots,m. \tag{37}$$

$$\lambda_i\geqslant 0, i=0,1,\cdots,m. \blacksquare$$

上面定理中的λ_i称为Lagrange乘子，$\lambda_i g_i(x^*)=0, i=1,\cdots,m$称为互补松弛条件．

例4 研究R^2中的(NLP_1)：

$$\begin{aligned}
\min \quad & f(x)=(\xi_1-3)^2+(\xi_2-2)^2 \\
\text{s.t} \quad & g_1(x)=\xi_1^2+\xi_2^2\leqslant 5, \\
& g_2(x)=\xi_1+2\xi_2\leqslant 4, \\
& g_3(x)=-\xi_1\leqslant 0, \\
& g_4(x)=-\xi_2\leqslant 0, x=(\xi_1,\xi_2).
\end{aligned}$$

图 7 是其可行域. 极小值点是$x^*=(2,1)$, $I(x^*)=\{1,2\}$. 计算知

$$\nabla f(x^*)=(-2,-2),\nabla g_1(x^*)=(4,2),$$
$$\nabla g_2(x^*)=(1,2).$$

令$\lambda_0=3,\lambda_1=1,\lambda_2=2,\lambda_3=\lambda_4=0$,可以验证(32),(33)式成立,即在$x^*$满足Fritz John条件.

再考察点$\bar{x}=(0,0)$, $I(\bar{x})=\{3,4\}$. 计算知

$$\nabla f(\bar{x})=(-6,-4),\nabla g_3(\bar{x})=(-1,0),$$
$$\nabla g_4(\bar{x})=(0,-1).$$

令$\lambda_1=\lambda_2=0$. 要使

$$\lambda_0(-6,-4)+\lambda_3(-1,0)+\lambda_4(0,-1)=(0,0)$$

成立,当且仅当$\lambda_3=-6\lambda_0$, $\lambda_4=-4\lambda_0$. 如果$\lambda_0>0$,则 λ_3, λ_4 小于0,,与λ_i的非负性矛盾;如果$\lambda_0=0$,则$\lambda_3=\lambda_4=0$与λ_i不全为0矛盾. 故$\bar{x}$不满足Fritz John条件,它不是极小值点.

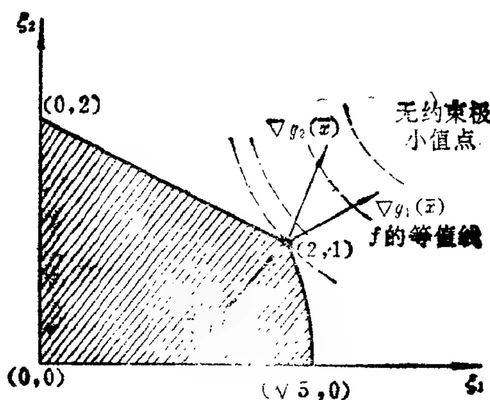


图 7

例5 再研究例 2

现在证明在局部极小值点$x^*=(1,0)$, 上述 Fritz John 条

件成立，$I(x^*)=\{1,3\}$.

$\nabla f(x^*)=(-1,0)$，$\nabla g_1(x^*)=(0,1)$，$\nabla g_3(x^*)=(0,-1)$.

要使

$$\lambda_0(-1,0)+\lambda_1(0,1)+\lambda_3(0,-1)=\theta$$

成立，只有 $\lambda_0=0$，这时 $\lambda_1=\lambda_3=\alpha$，$\alpha$ 是任意正数.

象定理 2.8 后面所说的情形一样，也存在平凡满足 Fritz John 条件的非局部极小值点．如果一点 $\bar{x}$ 满足$\nabla f(\bar{x})=\theta$ 或对某个 $i\in I(\bar{x})$满足$\nabla g_i(\bar{x})=\theta$，这时可以令相应的 Lagrange乘子为正数，其余乘子均为 0，从而在 $\bar{x}$ 满足 Fritz John 条件．同样，如果只存在等式约束，例如$g(x)=0$．用 $g_1(x)=g(x)\leqslant 0$，$g_2(x)=-g(x)\leqslant 0$ 代替 $g(x)=0$．令 $\lambda_1=\lambda_2=\alpha$，$\alpha$ 是任意正数，而其他乘子均为 0．于是在每个可行点，Fritz John 条件成立.

如果 Lagrange 乘子 $\lambda_0=0$，则 Fritz John 条件就不能利用关于目标函数的梯度的任何信息．所以当 $\lambda_0=0$时，Fritz John 条件在确定极小值点位置方面没有实用价值，因而 $\lambda_0>0$ 的情形才更有意义．为了保证 $\lambda_0>0$，可以对约束加上一定的条件，这种条件通常称为约束规格.

在上面的例 4 中，$\lambda_0>0$，例 5 中，$\lambda_0=0$．它们的差别在于例 4 中的紧约束的梯度线性无关，而例 5 中的紧约束的梯度线性相关．从下面的定理可知，紧约束的梯度线性无关的确可以保证 $\lambda_0>0$.

定理 2.10（(NLP$_1$)的 Kuhn-Tucker 必要条件）如果 x^* 是(NLP$_1$)的局部极小值点，$i\in I(x^*)$时，$\nabla g_i(x^*)$线性无关，则存在非负的实数 λ_i，$i=1,\cdots m$，使

$$\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)=\theta, \tag{38}$$

$$\lambda_i g_i(x^*)=0, \tag{39}$$

$$\lambda_i \geqslant 0,\ i=1,\cdots,m.$$

证明 设 x^*是局部极小值点，由定理 2.9，存在不全为0的非负实数 $\bar{\lambda}_i$，$i=0,1,\cdots m$，使

$$\bar{\lambda}_0 \nabla f(x^*) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(x^*) = \theta, \tag{40}$$

$$\bar{\lambda}_i g_i(x^*) = 0. \tag{41}$$

$\bar{\lambda}_0$一定不为0，因为否则 $\bar{\lambda}_0=0$，由(40),(41)式知

$$\sum_{i\in I(x^*)} \bar{\lambda}_i \nabla g_i(x^*) = \theta,\ \bar{\lambda}_i \text{不全为}0,$$

这说明 $\nabla g_i(x^*)$线性相关，其中 $i\in I(x^*)$这与定理的条件矛盾. 设 $\lambda_i=\bar{\lambda}_i/\lambda_0$，由(40),(41)式得

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) = \theta,$$

$$\lambda_i g_i(x^*) = 0,$$

$$\lambda_i \geqslant 0,\ i=1,\cdots,m. \blacksquare$$

可以验证，在例 4 中的 $x^*=(2.1)$，Kuhn-Tucker 必要条件中相应的$\lambda_1=\frac{1}{3}$，$\lambda_2=\frac{2}{3}$，$\lambda_3=\lambda_4=0$.

Kuhn-Tucker 条件说明，在局部极小值点 x^*，有

$$-\nabla f(x^*) \in \text{cone}\{\nabla g_i(x^*) \mid i\in I(x^*)\} \tag{42}$$

在图 8 中，x_1就是满足上述关系的点，而 x_2则不满足.

仍称定理 2.10 中的 λ_i 为 Lagrange 乘子. 称(39)为互补松弛条件

对于(NLP_2)，有相应的结论成立. 下面只叙述结果而不证明. 有兴趣的读者可以参见[40].

定理 2.8' 设 x^*是(NLP_2)的局部极小值点. $\nabla h_j(x^*)$线性无关，$j=1,\cdots,p$，则 $G(x^*)\cap F(x^*)\cap H(x^*)=\varnothing$，其中

$$G(x^*)=\{y \mid \langle y,\ \nabla g_i(x^*)\rangle < 0, i\in I(x^*)\},$$

$$F(x^*)=\{y \mid \langle y, \nabla f(x^*)\rangle<0\},$$
$$H(x^*)=\{y \mid \langle y, \nabla h_j(x^*)\rangle=0, \ j=1,\cdots,p\}.$$

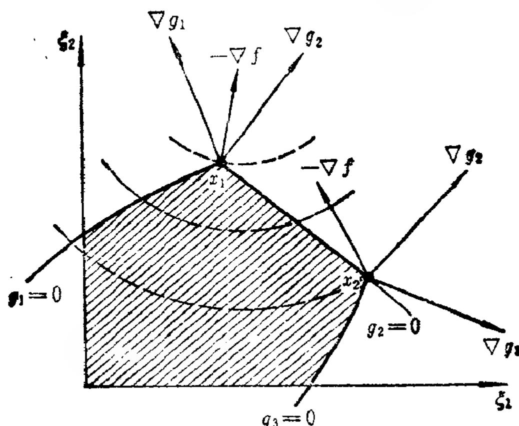


图 8 Kunn-Tucker 条件的几何说明

定理 2.9′ 设 x^* 是(NLP_2)的局部极小值点，则存在不全为 0 的实数$\lambda_1, \cdots, \lambda_m, \mu_1, \cdots, \mu_p$，满足

$$\lambda_0 \nabla f(x^*)+\sum_{i=1}^{m} \lambda_i \nabla g_i(x^*)+\sum_{j=1}^{p} \mu_j \nabla h_j(x^*)=\theta, \quad (43)$$

$$\lambda_i g_i(x^*)=0, \quad (44)$$

$$\lambda_0, \lambda_i \geqslant 0, \ i=1,\cdots,m.$$

上面的定理仍称为 Fritz John 条件，λ_i, μ_j 仍称为 Lagrange 乘子.

例 6 研究 R^2 中的(NLP_2)：

$$\min f(x)=(\xi_1-3)^2+(\xi_2-2)^2$$
$$\text{s.t } g_1(x)=\xi_1^2+\xi_2^2-5\leqslant 0$$
$$g_2(x)=-\xi_1\leqslant 0$$
$$g_3(x)=-\xi_2\leqslant 0$$
$$h_1(x)=\xi_1+2\xi_2-4=0, x=(\xi_1,\xi_2).$$

这是例 4 中不等式约束 $g_2(x)\leqslant 0$ 变为 $g_2(x)=0$ 而成的. 局部极小值点 $x^*=(2,1)$, $I(x^*)=\{1\}$. 对于

$$\lambda_0(-2,-2)-\lambda_1(4.2)-\mu_1(1,2)=(0,0),$$

只需取 $\lambda_0=3$, $\lambda_1=1$, $\mu_1=2$即可, 所以在 x^*, 定理 2.9′的结论是成立的.

例7 研究 R^2 中的(NLP_2):

$$\min f(x)=-\xi_1$$

$$\text{s.t.}\ h_1(x)=\xi_2-(1-\xi_1)^3=0$$

$$h_2(x)=-\xi_2-(1-\xi_1)^3=0, x=(\xi_1,\xi_2).$$

从图 9 知, 这个问题只有一个可行点 $x^*=(1,0)$. 计算知

$$\nabla f(x^*)=(-1,0),\ \nabla h_1(x^*)=(0,1),\ \nabla h_2(x^*)=(0,-1)$$

要

$$\lambda_0(-1,0)+\mu_1(0,1)+\mu_2(0,-1)=(0,0)$$

成立, 只有 $\lambda_0=0$, $\mu_1=\mu_2=\alpha$, 其中 α 为任意实数, 即 Fritz John 必要条件在 x^* 成立.

正如定理 2.9 后所作的说明一样, 如果 $\lambda_0=0$,则 Fritz John 条件在确定局部极小值点位置时没有实用价值. 为了保证$\lambda_0>0$, 仍需对约束加上一些限制, 即约束规格.

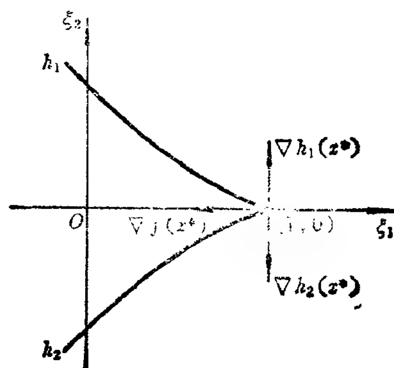


图 9

定理 2.10′ 设 x^* 是(NLP_2)的局部极小值点，且$\nabla g_i(x^*)$，$i\in I(x^*)$，$\nabla h_j(x^*)$，$j=1,\cdots,p$，线性无关，则存在实数λ_i，$i=1,\cdots,m$，μ_j，$j=1,\cdots,p$，使

$$\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)-\sum_{j=1}^{p}\mu_j\nabla h_j(x^*)=\theta, \tag{45}$$

$$\lambda_i g_i(x^*)=0, \tag{46}$$

$$\lambda_i\geqslant 0.\quad i=1,\cdots,m.$$

定理 2.10*的条件仍称为 Kuhn-Tucker 条件，λ_i，μ_j 仍称为 Lagrange 乘子.

在例 6 中的 $x^*=(2,1)$取$\lambda_1=\frac{1}{3}$，$\lambda_2=\lambda_3=0$. $\mu_1=\frac{2}{3}$即可，而例 7 的 $x^*=(1,0)$不满足定理 2.10*的条件，因为$\nabla h_1(x^*)$与$\nabla h_2(x^*)$线性相关.

3. 约束规格

前面我们是在 $\nabla g_i(x^*)$，$i\in I(x^*)$线性无关的约束规格下得到 Kuhn-Tucker 必要条件的. 事实上，还有其他约束规格保证这个必要条件的成立. 下面讨论一些约束规格并说明它们之间的关系.

定义 2.5 设 $X\subset R^n$，$\bar{x}\in \mathrm{cl}X$，称集合

$T(\bar{x})=\{y\mid y=\lim\limits_{k\to\infty}\lambda_k(x_k-\bar{x})$，$\lambda_k>0$，$x_k\in X$，$x_k\to\bar{x}\}$是 X 在 $\bar{x}$ 的正切锥.

根据定义，如果存在一个收敛于 $\bar{x}$ 的可行点列$\{x_k\}$，使弦 $x_k-\bar{x}$ 的方向收敛于 y，则 y 属于 $T(\bar{x})$. 可以证明，正切锥是闭锥.

仍先考虑本节所研究的形式(1)的极值问题，即

$$\min f(x)$$
$$\text{s.t } x\in S.$$

的问题.

定理 2.11 设 $S\subset R^n$，$S\neq\varnothing$. x^* 是(1)的局部极小值点，

$f(x)$在 x^* 可微，则 $T(x^*)\cap F(x^*)=\varnothing$

证明 设 $y\in T(x^*)$，即 $y=\lim\limits_{k\to\infty}\lambda_k(x_k-x^*)$，其中 $\forall k$，$\lambda_k>0$，$x_k\in S$，$x_k\to x^*$. 根据 $f(x)$在 x^* 的可微性，得

$$f(x_k)-f(x^*)=\langle\nabla f(x^*),\ x_k-x^*\rangle+o(|x_k-x^*|). \tag{47}$$

因为 x^* 是局部极小值点，当 k 充分大时，$f(x_k)\geqslant f(x^*)$. 故由(43)得

$$\langle\nabla f(x^*),\ x_k-x^*\rangle+o(|x_k-x^*|)\geqslant 0.$$

乘以 λ_k，令 $k\to\infty$，则有

$$\langle\nabla f(x^*),\ y\rangle\geqslant 0.$$

而 $F(x^*)=\{y\,|\,\langle\nabla f(x^*),\ y\rangle<0\}$，所以$F(x^*)\cap T(x^*)=\varnothing$. ▌

设 $G'(x)=\{y\,|\,\langle\nabla g_i(x),\ y\rangle\leqslant 0,\ i\in I(x)\}$.

定理 2.12 设 x^* 是(NLP$_1$)的局部极小值点，且 $T(x^*)=G'(x^*)$，则存在非负实数 $\lambda_i,i=1,\cdots,m$，使

$$\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)=\theta, \tag{48}$$

$$\lambda_i g_i(x^*)=0. \tag{49}$$

证明 由定理 2.11，$F(x^*)\cap T(x^*)=\varnothing$. 故由条件知，$F(x^*)\cap G'(x^*)=\varnothing$. 换言之，下面的不等式组

$\langle\nabla f(x^*),\ y\rangle<0$，$\langle\nabla g_i(x^*),\ y\rangle\leqslant 0, i\in I(x^*)$无解.

故由定理 1.9，得到存在非负实数 λ_i，$i\in I(x^*)$，有

$$\nabla f(x^*)+\sum_{i\in I(x^*)}\lambda_i\nabla g_i(x^*)=\theta. \tag{50}$$

如果在 $i\notin I(x^*)$时，令 $\lambda_i=0$，便得到(48)、(49)式. ▌

$T(x)=G'(x)$ 称为 Abadie 约束规格.

为讨论其他约束规格，引入下面的概念.

定义 2.6 设 $S\subset R^n$. 对于非零向量 y，如果存在$\delta>0$和从 R 到 R^n 的一元向量值函数 $\alpha(\lambda)$，使得 $\lambda\in(0,\delta)$时，$\alpha(\lambda)\in S$，

$\alpha(0)=\bar{x}$，且

$$\lim_{\lambda\downarrow 0+}\frac{\alpha(\lambda)-\alpha(0)}{\lambda}=y, \tag{51}$$

则称 y 是 S 在 $\bar{x}$ 的一个可达方向．S 在 $\bar{x}$ 的全体可达方向的集合，称为 S 在 $\bar{x}$ 的可达方向锥，用 $A(\bar{x})$ 表示．

定理 2.13 在(NLP_1)中，X 是可行解集，$\bar{x}\in X$，则

$$\mathrm{cl}G(\bar{x})\subset \mathrm{cl}D(\bar{x})\subset \mathrm{cl}A(\bar{x})\subset T(\bar{x})\subset G'(\bar{x}). \tag{52}$$

证明 容易证明，$D(\bar{x})\subset A(\bar{x})\subset T(\bar{x})\subset G'(\bar{x})$．但因为 $T(\bar{x})$ 是闭的．所以 $\mathrm{cl}D(\bar{x})\subset \mathrm{cl}A(\bar{x})\subset T(\bar{x})\subset G'(\bar{x})$．由定理2.8，$G(\bar{x})\subset D(\bar{x})$．所以(52)式成立．▌

下面是保证 Kuhn-Tucker 必要条件成立的约束规格．

1）线性无关的约束规格．

集合 S 是开集，$g_i(x)$ 在 $i\in I(\bar{x})$ 时在 $\bar{x}$ 可微，$i\notin I(\bar{x})$ 时在 $\bar{x}$ 连续，$i\in I(\bar{x})$ 时 $\nabla g_i(x)$ 线性无关．

2）Cottle 约束规格．

集合 S 是开集，$i\in I(\bar{x})$ 时 $g_i(x)$ 在 $\bar{x}$ 连续，且

$$\mathrm{cl}G(\bar{x})=G'(\bar{x}). \tag{53}$$

3）Zangwill 约束规格．

$$\mathrm{cl}D(\bar{x})=G'(\bar{x}). \tag{54}$$

4）Kuhn-Tucker 约束规格．

$$\mathrm{cl}A(\bar{x})=G'(\bar{x}). \tag{55}$$

这些约束规格的关系是：

1）线性无关约束规格成立，则 Cottle 约束规格一定成立．因为这时 $\sum_{i\in I(\bar{x})}\lambda_i\nabla g_i(\bar{x})=0$ 没有非零解，故由定理 1.15 知存在 y，使 $\langle\nabla g_i(\bar{x}),y\rangle<0$，$i\in I(\bar{x})$，于是知 $G(\bar{x})\neq\varnothing$，容易验证，$\mathrm{cl}G(\bar{x})=G'(\bar{x})$．

2）Cottle 约束规格成立，则 Zangwill 约束规格一定成立，

这是定理 2.13 的结果.

根据同样的定理，有

3）Zangwill 约束规格成立时，Kuhn-Tucker 约束规格成立；Kuhn-Tucker 约束规格成立时，Abadie 约束规格成立.

根据定理 2.12，Abadie 约束规格成立使 Kuhn-Tucker 必要条件成立. 于是可知，所列约束规格都能保证 Kuhn-Tucker必要条件的成立.

4．二阶条件

下面讨论与二阶导数有关的条件.

设

$Z_1(x)=\{y|\langle y,\ \nabla g_i(x)\rangle=0,\ i\in I(x),\ \langle y,\ \nabla h_j(x)\rangle=0,\ j=1,\cdots,p\}$.

定义 2.7 如果每一个非零的 $y\in Z_1(\bar{x})$ 与包含在 S 的边界中的二次可微的弧相切，则称在 $\bar{x}$ 二阶约束规格成立. 换言之，对于每一个 $y\in Z_1(x)$，存在一个定义在 $[0,\ \varepsilon]\subset R$，值域为 R^n 中的向量的二次可微向量函数 α，满足 $\alpha(0)=\bar{x}$，$g_i(\alpha(\rho))=0$，$i\in I(\bar{x})$，$h_j(\alpha(\rho))=0$，$j=1,\cdots,p$，其中 $0\leqslant\rho\leqslant\varepsilon$. 且对于某个正数 λ，有

$$\frac{\mathrm{d}\alpha(0)}{\mathrm{d}\rho}=\lambda y. \tag{56}$$

定理 2.14 （二阶必要条件）设 x^* 是(NLP_2)的局部极小值点，存在 $\lambda_1,\cdots,\lambda_m$，$\mu_1,\cdots,\mu_p$，使

$$\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)+\sum_{j=1}^{p}\mu_j\nabla h_j(x^*)=\theta, \tag{57}$$

$$\lambda_i g_i(x^*)=0,\ i=1,\cdots,m, \tag{58}$$
$$\lambda_i\geqslant 0.$$

又设二阶约束规格在 x^* 成立，则对于非零的 $y\in Z_1(x^*)$ 满足

$$\langle y, [f''(x^*) + \sum_{i=1}^{m} \lambda_i g_i''(x^*) + \sum_{j=1}^{p} \mu_j h''_j(x^*)] y \rangle \geq 0 \tag{59}$$

证明 设 $y \neq \theta$, $y \in Z_1(x^*)$, $\alpha(\rho)$是二阶约束规格中所设的向量值函数，即 $\alpha(0) = x^*$, $\dfrac{d\alpha(0)}{d\rho} = y$（不妨设 $\lambda = 1$）. 设$\dfrac{d^2\alpha(0)}{d\rho^2} = w$，故有

$$\frac{d^2 g_i(\alpha(0))}{d\rho^2} = \langle y, g_i''(x^*) y \rangle + \langle w, \nabla g_i(x^*) \rangle = 0 \quad i \in I(x^*), \tag{60}$$

$$\frac{d^2 h_j(\alpha(0))}{d\rho^2} = \langle y, h_j''(x^*) y \rangle + \langle w, \nabla h_j(x^*) \rangle = 0, j = 1, \cdots, p. \tag{61}$$

根据 (57)-(58)式及 $Z_1(x^*)$的定义，有

$$\frac{df(\alpha(0))}{d\rho} = \langle y, \nabla f(x^*) \rangle = \langle y, [-\sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) - \sum_{j=1}^{p} \mu_j \nabla h_j(x^*)] \rangle = 0.$$

因为 x^*是局部极小值点，$\dfrac{df(\alpha(0))}{d\rho} = 0$. 故$\dfrac{d^2 f(\alpha(0))}{d\rho^2} \geq 0$. 即

$$\frac{d^2 f(\alpha(0))}{d\rho^2} = \langle y, f''(x^*) y \rangle + \langle w, \nabla f(x^*) \rangle \geq 0. \tag{62}$$

将 (60), (61)式分别乘上相应的 λ_i, μ_j, 与(62)相加. 并注意到(57)-(58)式，得

$$\langle y, [f''(x^*) + \sum_{i=1}^{m} \lambda_i g_i''(x^*) + \sum_{j=1}^{p} \mu_j h_j''(x^*)] y \rangle \geq 0,$$

即 (59) 式成立. ▌

具体检验二阶约束规格的成立是很困难的. 可以证明.如果

$\nabla g_i(x), i\in I(x), \nabla h_j(x), j=1,\cdots,p$ 线性无关，则二阶约束规格也是成立的[41]．

为了研究充分条件，设 $I_1(x^*)$ 是满足 $g_i(x^*)=0$，且在(57)-(58)式中 $\lambda_i>0$ 的指标 i 的集合．显然 $I_1(x^*)\subset I(x^*)$．又设

$$Z_2(x)=\{y\mid\langle y,\nabla g_i(x)\rangle=0, i\in I_1(x^*), \langle y,\nabla g_i(x^*)\rangle\leqslant 0,$$
$$i\in I(x^*), \langle y,\nabla h_j(x^*)\rangle=0, j=1,\cdots,p\}.$$

定理 2.15 令 x^* 是(NLP_2)的可行点．如果存在 $\lambda_1,\cdots,\lambda_m,\mu_1,\cdots,\mu_p$ 满足

$$\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)-\sum_{j=1}^{p}\nabla h_j(x^*)=\theta, \tag{63}$$

$$\lambda_i g_i(x^*)=0, \qquad i=1,\cdots,m, \tag{64}$$

$$\lambda_i\geqslant 0.$$

且对每一个非零的 $y\in Z_2(x^*)$，有

$$\langle y,[f''(x^*)-\sum_{i=1}^{m}\lambda_i g_i''(x^*)-\sum_{j=1}^{p}\mu_j h_j''(x^*)]y\rangle>0, \tag{65}$$

则 x^* 是(NLP_2)的局部严格极小值点．

证明 设 x^* 不是局部严格极小值点，则存在一个可行点列 $\{x_k\}, x_k\neq x^*, \lim\limits_{k\to\infty}x_k=x^*$，且

$$f(x^*)\geqslant f(x_k). \tag{66}$$

设 $x_k=x^*+\beta_k y_k$，其中 $\beta_k>0, \|y_k\|=1$，不失一般性，可以认为序列 $\{\beta_k,y_k\}\xrightarrow[k\to\infty]{}\{0,\bar{y}\}, \|\bar{y}\|=1$，于是有

$$g_i(x_k)-g_i(x^*)=\beta_k\langle y_k,\nabla g_i(x^*+\eta_{ik}\beta_k y_k)\rangle\leqslant 0, i\in I(x^*). \tag{67}$$

$$h_j(x_k)-h_j(x^*)=\beta_k\langle y_k,\nabla h_j(x^*+\bar{\eta}_{jk}\beta_k y_k)\rangle=0,$$
$$j=1,\cdots,p. \tag{68}$$

$$f(x_k)-f(x^*)=\beta_k\langle y_k,\nabla f(x^*+\eta_k\beta_k y_k)\rangle\leqslant 0. \tag{69}$$

其中 $0<\eta_k,\eta_{ik},\bar{\eta}_{jk}<1$．用 β_k 除(67)-(69)式，令 $k\to\infty$ 取极限，得

$$\langle \bar{y}, \nabla g_i(x^*) \rangle \leqslant 0, \qquad i \in I(x^*), \tag{70}$$

$$\langle \bar{y}, \nabla h_j(x^*) \rangle = 0, \tag{71}$$

$$\langle \bar{y}, \nabla f(x^*) \rangle \leqslant 0. \tag{72}$$

如果对一个 $i \in I_1(x^*)$,(70)式是严格不等式，结合(63),(70),(71)式得

$$\langle \bar{y}, \nabla f(x^*) \rangle = -\sum_{i=1}^{m} \lambda_i \langle \bar{y}, \nabla g_i(x^*) \rangle - \sum_{j=1}^{p} \mu_j \langle \bar{y}, \nabla h_j(x^*) \rangle > 0$$

这与 (72) 式矛盾. 故对全部 $i \in I_1(x^*)$ 及 $\bar{y} \in Z_2(x^*)$, $\langle y, \nabla g_i(x^*) \rangle = 0$.

由 Taylor 定理,有

$$g_i(x_k) = g_i(x^*) + \beta_k \langle y_k, \nabla g_i(x^*) \rangle + \frac{1}{2} \beta_k^2 \langle y_k, g_i''(x^* + \xi_{ik} \beta_k y_k) y_k \rangle \leqslant 0,$$
$$i = 1, \cdots, m. \tag{73}$$

$$h_j(x_k) = h_j(x^*) + \beta_k \langle y_k, \nabla h_j(x^*) \rangle + \frac{1}{2} \beta_k^2 \langle y_k, h_j''(x^* + \bar{\xi}_{jk} \beta_k y_k) y_k \rangle = 0,$$
$$j = 1, \cdots, p. \tag{74}$$

$$f(x_k) - f(x^*) = \beta_k \langle y_k, \nabla f(x^*) \rangle + \frac{1}{2} \beta_k^2 \langle y_k, f''(x^* + \xi_k \beta_k y_k) y_k \rangle \leqslant 0, \tag{75}$$

其中 $0 < \xi_{ik}, \xi_{jk}, \bar{\xi}_k < 1$.

将(73),(74)式分别乘相应的 λ_i、μ_j,再与(75)式相加,得

$$\beta_k \langle y_k, \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(x^*) \rangle$$

$$+\frac{1}{2}\beta_k^2\langle y_k,[f''(x^*+\xi_k\beta_k y_k)+\sum_{i=1}^{m}\lambda_i g_i''(x^*+\xi_{ik}\beta_k y_k)$$

$$+\sum_{j=1}^{p}\mu_j h_j''(x^*+\bar{\xi}_{jk}\beta_k y_k)]y_k\rangle\leqslant 0. \tag{76}$$

由(63)式知(76)式左边第一项是0．将(76)式除以$\frac{1}{2}\beta_k^2$，令$k\to\infty$取极限．得

$$\langle\bar{y},[f''(x^*)+\sum_{i=1}^{m}\lambda_i g_i''(x^*)+\sum_{j=1}^{p}\mu_j h_j''(x^*)]\ \bar{y}\rangle\leqslant 0. \tag{77}$$

因为$\bar{y}$非零，且$\bar{y}\in Z_2(x^*)$．故与(65)式矛盾，所以x^*是局部严格极小值点．

§3 凸函数的极值与凸规划

本节研究形如

$$\min\ f(x) \tag{1}$$

$$\text{s.t.}\ x\in S$$

的约束极值问题，其中$f(x)$是R^n上的凸函数(不一定可微)，S是R^n中的凸集，这类极值问题称为凸规划(CP)．首先，我们讨论$S=R^n$的情形，再讨论(1)的最优性条件，最后对一类特殊的S，即由凸不等式所定义的S的情形讨论其最优性条件．

1．凸函数的极值

定理 3.1 设$f(x)$是R^n上的正常凸函数，则$f(x)$的每一个局部极小值也是全局极小值．

证明 设x^*是$f(x)$的局部极小值点，则存在$\delta>0$，当$x\in\{x\,|\,|x-x^*|<\delta\}$时，有

$$f(x)\geqslant f(x^*). \tag{2}$$

设 $\forall z\in R^n$，则当 λ 充分小时，$0<\lambda<1$，有

$$(1-\lambda)x^*+\lambda z\in\{x\,|\,|x-x^*|<\delta\}. \tag{3}$$

因为 $f(x)$ 是凸函数，故

$$f(x^*)\leqslant f[(1-\lambda)x^*+\lambda z]\leqslant(1-\lambda)f(x^*)+\lambda f(z). \tag{4}$$

由(4)式知 $f(x^*)\leqslant f(z)$，故 x^* 是全局极小值点. ▌

根据定理 3.1，以后对于凸函数就不再区别局部极小和全局极小，而只简单地称极小.

定义 3.1 设 $f(x)$ 是 R^n 上的凸函数，称达到 $\inf\limits_{x\in R^n} f(x)$ 的 x 的集合是 $f(x)$ 的极小集.

定理 3.2 $f(x)$ 的极小集是凸集.

证明 设 $\alpha^*=\inf\limits_{x\in R^n} f(x)$，则 $f(x)$ 的极小集是水平集

$$\{x\,|\,f(x)\leqslant\alpha^*\},$$

由第二章定理 1.8，知极小集是凸集. ▌

定理 3.3 设 $f(x)$ 是 R^n 上的正常凸函数，则 x^* 属于 $f(x)$ 的极小集的充分必要条件是 $\theta\in\partial f(x^*)$.

证明 如果 $\theta\in\partial f(x^*)$，则对 $\forall x\in R^n$，有

$$f(x)-f(x^*)\geqslant\langle x-x^*,\theta\rangle, \tag{5}$$

即 $f(x)\geqslant f(x^*)$，x^* 属于 $f(x)$ 的极小集.

反之，设 x^* 属于 $f(x)$ 的极小集，即 $\forall x\in R^n$，$f(x)\geqslant f(x^*)$. 而这与对于 $\forall x\in R^n$，有

$$f(x)-f(x^*)\geqslant 0=\langle x-x^*,\theta\rangle,$$

等价，所以 $\theta\in\partial f(x^*)$. ▌

推论 3.3.1 设 $f(x)$ 是 R^n 上的正常凸函数，在 x^* 可微，则 x^* 属于 $f(x)$ 的极小集的充分必要条件是 $\nabla f(x^*)=\theta$.

证明留给读者.

定理 3.4 设 $f(x)$ 是 R^n 上的正常闭凸函数，则

1) $\inf\limits_{x\in R^n} f(x)=-f^*(\theta)$，因而 $f(x)$ 有下界的充分必要条件是 $\theta\in \mathrm{dom} f^*$.

2) $f(x)$ 的极小集是 $\partial f^*(\theta)$，因而 $f(x)$ 达到下确界的充分必要条件是 $f^*(x^*)$ 在 $x^*=\theta$ 次可微. 特别地，当 $\theta\in \mathrm{ri}\,(\mathrm{dom} f^*)$ 时这个条件是满足的.

3) $f(x)$ 的下确界有限但不能达到的充分必要条件是 $f^*(\theta)$ 有限且对某一个 y，$f^{*\prime}(\theta:y)=-\infty$.

4) $f(x)$ 的极小集非空有界的充分必要条件是 $\theta\in \mathrm{int}\,(\mathrm{dom} f^*)$

5) $f(x)$ 的极小集仅含唯一的向量 x 的充分必要条件是 $f^*(x^*)$ 在 $x^*=\theta$ 可微且 $x=\nabla f^*(\theta)$.

证明 1) 由共轭函数的定义，有

$$f^*(x^*)=\sup_{x\in R^n}\{\langle x, x^*\rangle-f(x)\}. \tag{6}$$

故 $f^*(\theta)=\sup\limits_{x\in R^n}\{-f(x)\}=-\inf\limits_{x\in R^n} f(x)$.

2) 由定理 3.3，x_0 属于 $f(x)$ 的极小集的等价条件是 $\theta\in \partial f(x_0)$. 这又等价于 $x^*=\theta$ 是 $f(x)$ 在 x_0 的次梯度，根据第三章定理 1.9 的 1) 和 1*)，知 $x_0\in\partial f^*(\theta)$，即 $f(x)$ 的极小集是 $\partial f^*(\theta)$.

3)，4)，5) 的证明留给读者. ▌

2. 约束是凸集的凸规划

现在回到(1) 式的问题.

定理 3.5 设 $f(x)$ 是 R^n 上的正常凸函数，S 是 R^n 中的凸集，且存在 $x_1\in S$，$f(x)$ 在 x_1 连续，则 x^* 是 $f(x)$ 在 S 上的极小值点的充分必要条件是

$$\partial f(x^*)\cap K_S^*(x^*)\neq\varnothing, \tag{7}$$

其中 $K_S(x^*)=\mathrm{cone}(S-x^*)=\{\bar{x}\mid\lambda>0$ 充分小时 $x^*+\lambda\bar{x}\in S\}$.

证明 设

$$\delta(x|S)=\begin{cases}0 & x\in S,\\ +\infty & x\notin S.\end{cases}$$

故求 $f(x)$ 在 S 上的极小值等价于求 $F(x)=f(x)+\delta(x|S)$ 在 R^n 上的极小值. 由第三章定理 4.2,有

$$\partial F(x)=\partial f(x)+\partial\delta(x|S). \tag{8}$$

但 $\partial\delta(x|S)=-K_S^*(x)$, 故 $\partial F(x)=\partial f(x)-K_S^*(x)$. 由定理 3.3, x^* 属于 $F(x)$ 的极小集的等价条件是

$$\theta\in\partial f(x^*)-K_S^*(x^*),$$

这正好与(7) 式等价. ▌

考虑到共轭锥的定义,定理 3.5 的另一形式是:

定理 3.5′ 设 $f(x)$ 是 R^n 上的正常凸函数, S 是 R^n 中的凸集,则 x^* 是 $f(x)$ 在 S 的极小值点的充分必要条件是存在 $y\in\partial f(x^*)$,满足

$$\langle y,\quad x-x^*\rangle\geqslant 0,\qquad \forall x\in S. \tag{9}$$

定理 3.5 说明了(1) 的最优解的重要特征. 特别地,当 $f(x)$ 可微、S 是开集时,这个特征就是导数为 0 的条件. 同时它也表明,如果 x_0 不是最优解时,应有一个 $x\in S$,使

$$\langle\nabla f(x_0),x-x_0\rangle<0,$$

因而存在一个改进解的明显途径,即从 x_0 沿 $x-x_0$ 的方向前进,改进的步长大小由下面的一维子问题决定:

$$\begin{aligned}\min\quad & f(x_0+\lambda(x-x_0))\\ \text{s.t}\quad & x_0+\lambda(x-x_0)\in S,\\ & \lambda\geqslant 0.\end{aligned} \tag{10}$$

这就是最优化方法中可行方向法的根据.

例 1 研究凸规划

$$\min\ f(\xi_1,\xi_2)=\left(\xi_1-\frac{3}{2}\right)^2+(\xi_2-5)^2$$

$$\text{s.t}\quad -\xi_1+\xi_2\leqslant 2,$$

$$2\xi_1+3\xi_2\leqslant 11$$
$$-\xi_1\leqslant 0$$
$$-\xi_2\leqslant 0.$$

$f(\xi_1,\xi_2)$表示与点$\left(\frac{3}{2},5\right)$的距离的平方. 约束是由线性不等式组决定的多面体 S. 如图 10.

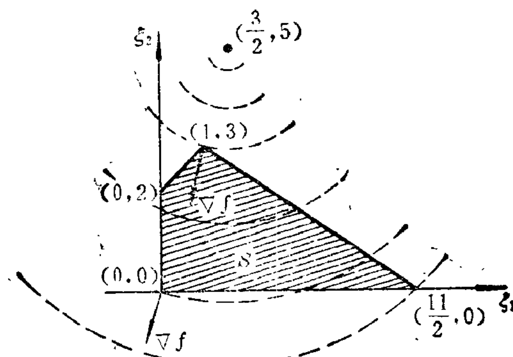


图 10

由图 10 知，最优解为 $x^*=(1,3)$. $\nabla f(1.3)=(-1,-4)$. 从几何上看,向量$(-1,-4)$与形式为(ξ_1-1,ξ_2-3)的每个向量的夹角均不超过$\frac{\pi}{2}$. 其中$(\xi_1,\xi_2)\in S$. 这说明满足定理3.5′的条件.

再考虑 $x_0=(0,0)$. 注意 $\nabla f(0,0)=(-3,-10)$,故对每个非θ的$x=(\xi_1,\xi_2)$, $-3\xi_1-10\xi_2<0$. 因此 x_0 不是最优解,所以可以从原点出发,沿 $x-\theta$ 的方向改进 $f(x)$的值. 当然,最好的局部方向是$-\nabla f(0,0)$,即$(3,10)$.

定理 3.6 设定理 3.5 中的条件成立,$S=S_1\cap\cdots\cap S_k$. 其中 S_i 是凸集,$i=1,\cdots,k$. $K_{S_i}(x^*)$的定义同定理 3.5, 且

$$\mathrm{int}S_1\cap\cdots\cap\mathrm{int}S_{k-1}\cap S_k\neq\varnothing,$$

则 x^* 是 $f(x)$ 在 S 上的极小值点的充分必要条件是存在 $x_i^*\in K_{S_i}^*(x^*)$, $i=1,\cdots,k$, $\bar{x}\in\partial f(x^*)$,满足

$$\bar{x}=x_1^*+\cdots+x_k^*. \tag{11}$$

证明 由第一章定理 8.16、定理 8.17,知

$$K_S(x^*)=\bigcap_{i=1}^{k} K_{S_i}(x^*). \tag{12}$$

如果 $x_0\in \operatorname{int} S_i, i=1,\cdots,k-1, x_0\in S_k$,则

$$\bar{x}_0=x_0-x^*\in \operatorname{int} K_{S_i}(x^*),\quad i=1,\cdots,k-1,$$

$$\bar{x}_0\in K_{S_k}(x^*).$$

故 $\operatorname{int} K_{S_1}(x^*)\cap\cdots\cap \operatorname{int} K_{S_{k-1}}(x^*)\cap K_{S_k}(x^*)\neq\varnothing.$ (13)

故由第一章定理 8.12 ,得

$$K_S^*(x^*)=K_{S_1}^*(x^*)+\cdots+K_{S_k}^*(x^*). \tag{14}$$

根据定理 3.5 ,存在 $\bar{x}\in\partial f(x^*)\cap K_S^*(x^*)$,因而由(14)式有

$$\bar{x}=x_1^*+\cdots+x_k^*, \tag{15}$$

其中 $x_i^*\in K_{S_i}^*(x^*), i=1,\cdots,k$. ▎

定理 3.7 设 $f(x)$是 R^n 上的正常凸函数,$S=S_1\cap\cdots\cap S_k$,其中 S_i 是凸集,$i=1,\cdots,k$.又存在 $x_1\in S$,$f(x)$ 在 x_1 连续.则 x^* 是 $f(x)$在 S 上的极小值点的必要条件是存在向量 $\bar{x}\in\partial f(x^*)$,$x_i^*\in K_{S_i}^*(x^*)$及等于 1 或 0 的 λ,满足

$$\lambda\bar{x}=x_1^*+\cdots+x_k^*. \tag{16}$$

如果 $\lambda=0$,则 $x_1^*,\cdots,x_k^*$ 中至少有一个非零,如果 $\lambda=1$,则上述条件也是充分的.

证明 设 x^* 是 $f(x)$在 S 上的极小值点,则由定理 3.5,存在 $\bar{x}\in\partial f(x^*)\cap K_S^*(x^*)$. 因为(10)成立,则由第一章定理 3.13,这时有两种可能.

1) (14) 成立,即(15) 成立. 这是(16) 中 $\lambda=1$ 的情形. 因为定理 3.6 给出的是充分必要条件,所以这时(16)也是充分条件.

2) 存在不全为 θ 的 $x_i^*\in k_{S_i}^*(x^*)$,满足

$$x^*_1+\cdots+x^*_k=\theta,$$

此即(16)中 $\lambda=0$ 的情形. ∎

如果(1) 中的"min"换为"max",就成为:

$$\max \quad f(x) \tag{17}$$
$$\text{s.t} \quad x\in S.$$

下面讨论(17)的局部极大值点的必要条件. 这个条件不是充分的,而且也不存在寻求较好解的局部信息. 所以,求凸函数的极大值比求凸函数的极小值要困难得多.

定理 3.8 设 $f(x)$ 是 R^n 上的凸函数, S 是 R^n 中的凸集,如果 x^* 是 $f(x)$ 在 S 上的局部极大值点,则

$$\langle y,x-x^*\rangle\leq 0,\forall x\in S, \tag{18}$$

其中 y 是 $f(x)$ 在 x^* 的任一次梯度.

证明 设 x^* 是局部极大值点,则存在 $\delta>0$,使

$$f(x)\leq f(x^*),\forall x\in S\cap\{x|\|x-x^*\|<\delta\}.$$

设 $x\in S$,当 $\lambda>0$ 充分小时, $x^*+\lambda(x-x^*)\in S\cap\{x|\|x-x^*\|<\delta\}$. 故

$$f[x^*+\lambda(x-x^*)]\leq f(x^*). \tag{19}$$

设 $y\in\partial f(x^*)$,结合(19)式得

$$0\geq f[x^*+\lambda(x-x^*)]-f(x^*)\geq\lambda\langle y,x-x^*\rangle. \tag{20}$$

因为 $\lambda>0$,于是 $\langle y,x-x^*\rangle\leq 0$. ∎

推论 3.8.1 设 $f(x)$ 是 R^n 上的凸函数, S 是 R^n 中的凸集如果 x^* 是 $f(x)$ 在 S 上的局部极大值点,且 $f(x)$ 在 x^* 可微,则

$$\langle\nabla f(x^*),x-x^*\rangle\leqslant 0,\forall x\in S. \tag{21}$$

证明留给读者.

上述条件仅仅是必要条件而不是充分条件.

例 2 研究 R 上的极大问题

$$\max \quad f(x)=x^2$$

$$\text{s.t}\quad x\in S=\{x\mid -1\leqslant x\leqslant 2\}$$

显然，$f(x)$在S上的极大值为4，在$x=2$处达到，而在$\bar{x}=0$，$\nabla f(\bar{x})=\theta$，因而对任意$x\in S$，均有$\langle\nabla f(\bar{x}),(x-\bar{x})\rangle=0$.但$\bar{x}=0$连局部极大值点都不是.

在例1中，$x_0=(0,0)$，$x_1=\left(\frac{11}{2},0\right)$都是局部极大值点，且都满足定理3.8的必要条件．但是，我们在点x_0却不能得到如何从x_0到达全局极大值点$x_1=\left(\frac{11}{2},0\right)$的任何局部信息，即使对$x_1$，也没有方便的判别方法使我们知道它就是全局极大值点.

定理 3.9 在(17)中设S是一个多面体，则存在极大值点x^*，其中x^*是S的一个极点.

证明 因$f(x)$在S中连续，S是紧致集，则$f(x)$在一点$x^*\in S$达到极大值，如果x^*是S的极点，结果得证．否则，由第一章定理9.2知

$$x^*=\sum_{i=1}^{k}\lambda_i x_i,\quad \lambda_i>0,\ \sum_{i=1}^{k}\lambda_i=1, \tag{22}$$

其中x_i是S的极点，因为$f(x)$是凸函数，故

$$f(x^*)\leqslant\sum_{i=1}^{k}\lambda_i f(x_i) \tag{23}$$

但x^*是极大值点，$f(x^*)\geqslant f(x_i)$，$i=1,\cdots,k$．于是结合(23)得到

$$f(x^*)=f(x_i),\ i=1,\cdots k, \tag{24}$$

即极点$x_1,\cdots,x_k$是极大值点．▌

3．约束是不等式的凸规划

i) 凸不等式组的解

定理 3.10 设C是R^n中的凸集，$f_1(x),\cdots,f_m(x)$是R^n上满足$\mathrm{ri}C\subset\mathrm{dom}f_i$的正常凸函数，则下面两个结论恰有一个成立：

1）存在一个 $x\in C$，使

$$f_1(x)<0,\cdots,f_m(x)<0,$$

2）存在不全为 0 的非负实数 $\lambda_1,\cdots,\lambda_m$，使

$$\sum_{i=1}^{m}\lambda_i f_i(x)\geqslant 0,\forall x\in C.$$

证明 设 1)成立，对于满足 1)的 x 和任意 $\lambda_i\geqslant 0,i=1,\cdots,m$，且至少有一个 $\lambda_i\neq 0$，有

$$\sum_{i=1}^{m}\lambda_i f_i(x)<0,$$

故 2）不成立.

设 1）不成立，要证明 2)成立. 设 $C\neq\varnothing$，否则结论平凡. 令

$$C_1=\{z=(\xi_1,\cdots,\xi_m)\in R^m\mid 存在x\in C,\ f_i(x)<\xi_i,i=1,\cdots,m\}.$$

可以证明 $C_1\neq\varnothing$，C_1 是凸集（请读者自证），令非正真锥

$$C_2=\{z=(\xi_1,\cdots,\xi_m)\in R^m\mid \xi_i\leqslant 0,i=1,\cdots,m\}$$

因为 1)不成立，所以 $C_1\cap C_2=\varnothing$，根据第一章定理 5.3，C_1、C_2 可以分离. 则存在非零向量 $z^*=(\lambda_1,\cdots,\lambda_m)$和一个实数 α，满足

$$\alpha\leqslant\langle z^*,z\rangle=\lambda_1\xi_1+\cdots+\lambda_m\xi_m,\forall z\in C_1 \tag{25}$$

$$\alpha\geqslant\langle z^*,z\rangle=\lambda_1\xi_1+\cdots+\lambda_m\xi_m,\forall z\in C_2 \tag{26}$$

因为 C_2 是非正真锥，(26)式表明 $\alpha\geqslant 0$，由(25)式，$\lambda_i\geqslant 0$，$i=1,\cdots,m$，由(25)，总存在使 $f_i(x)<\xi_i,i=1,\cdots,m$的$x\in C$，有

$$\lambda_1\xi_1+\cdots+\lambda_m\xi_m\geqslant 0 \tag{27}$$

对于集合 $D=C\cap\mathrm{dom}f_1\cap\cdots\cap\mathrm{dom}f_m$ 中的每一个 x 及任意 $\varepsilon>0$，令

$$\xi_i=f_i(x)+\varepsilon,i=1,\cdots,m$$

则应有

$$\lambda_1[f_1(x)+\varepsilon]+\cdots+\lambda_m[f_m(x)+\varepsilon]\geqslant 0. \tag{28}$$

由 ε 的任意性，得

$$\lambda_1 f_1(x)+\cdots+\lambda_m f_m(x)\geqslant 0. \tag{29}$$

故凸函数 $f(x)=\lambda_1 f_1(x)+\cdots+\lambda_m f_m(x)$ 在 D 上非负、有限，由第二章推论 3.6.3，$f(x)$ 在 $\mathrm{cl}D$ 也非负，但 $\mathrm{ri}C\subset D$，故有

$$C\subset \mathrm{cl}(\mathrm{ri}C)\subset \mathrm{cl}D.$$

所以对于每一个 $x\in C$，$f(x)\geqslant 0$，故 2) 成立。▌

下面的例子说明定理 3.10 中 $\mathrm{ri}C\subset \mathrm{dom}f_i$ 的条件是必要的。

例 3 在 R 中，设

$$f_1(x)=\begin{cases}-x^{1/2} & x\geqslant 0,\\ +\infty & x<0.\end{cases}$$

$$f_2(x)=x,\quad C=R.$$

显然不存在 $x\in C$，满足 $f_1(x)<0$，$f_2(x)<0$。另一方面，仅当 $\lambda_1=\lambda_2=0$ 时，$\lambda_1 f_1(x)+\lambda_2 f_2(x)\geqslant 0$（实际上等号成立），$\forall x\in C$。故定理 3.10 中的 1)、2) 均不成立。可以检验，$\mathrm{ri}C\subset \mathrm{dom}f_1$ 这个条件不成立。

定理 3.11 设 C 是 R^n 中的凸集，$f_1(x),\cdots,f_k(x)$ 是 R^n 上的正常凸函数，$\mathrm{ri}C\subset \mathrm{dom}f_i$，$i=1,\cdots,k$，$f_{k+1}(x),\cdots,f_m(x)$ 是 R^n 上的仿射函数，不等式组

$$f_{k+1}(x)\leqslant 0,\cdots,f_m(x)\leqslant 0,$$

至少有一个解 $x\in \mathrm{ri}C$，则下面两个结论恰有一个成立：

1) 存在一个 $x\in C$，使

$$f_1(x)<0,\cdots,f_k(x)<0,f_{k+1}(x)\leqslant 0,\cdots,$$
$$f_m(x)\leqslant 0.$$

2) 存在非负实数 $\lambda_1,\cdots,\lambda_m$，$\lambda_1,\cdots,\lambda_k$ 中至少有一个非零，使

$$\sum_{i=1}^{m}\lambda_i f_i(x)\geqslant 0, \forall x\in C.$$

证明与定理 3.10 相似，略.

下面的定理可以看成是 Farkas 定理的推广.

定理 3.12 设 C 是 R^n 中的凸集，$f_0(x), f_1(x), \cdots, f_m(x)$ 是 R^n 上的凸函数，$\mathrm{ri}C\subset\mathrm{dom}f_i$. 如果不等式组

$$f_0(x)<0, f_i(x)\leqslant 0, i=1,\cdots,m, \tag{30}$$

在 C 中没有解，但存在一个 $x_0\in C$，满足

$$f_i(x_0)<0, \quad i=1,\cdots,m. \tag{31}$$

则下面两个结论恰有一个成立.

1) $f_0(x)\geqslant 0, \forall x\in C$

2) 存在不全为 0 的非负实数 $\lambda_1,\cdots,\lambda_m$，满足

$$f_0(x)-\sum_{i=1}^{m}\lambda_i f_i(x)\geqslant 0, \quad \forall x\in C.$$

证明 因为不等式组 (30) 在 C 中没有解，则由 $f_0(x)<0$, $f_1(x)<0,\cdots,f_m(x)<0$ 组成的不等式组在 C 中没有解，根据定理3.11，存在不全为 0 的非负实数 $\lambda_0', \lambda_1', \cdots, \lambda_m'$，满足

$$\lambda_0' f_0(x)+\sum_{i=1}^{m}\lambda'_i f_i(x)\geqslant 0, \forall x\in C. \tag{32}$$

不失一般，设 $\lambda_0'+\lambda_1'+\cdots+\lambda_m'=1$，如果 $\lambda_0'=1$，则 1) 成立；如果 $0<\lambda_0'<1$，用 λ_0'除(32)式两边，令 $\lambda_i=\lambda_i'/\lambda_0, i=1,\cdots,m$，则 2) 成立；如果 $\lambda_0'=0$，则由(32)式，有

$$\sum_{i=1}^{m}\lambda_i' f_i(x)\geqslant 0, \quad \forall x\in C. \tag{33}$$

因为 $x_0\in C$ 满足(31)式，故仅当 $\lambda_1'=\cdots=\lambda_m'=0$ 时(33)式才成立，这与 λ_i 不全为 0 矛盾，所以 $\lambda_0'=0$ 是不可能的. ▌

定理 3.13 设 C 是 R^n 中的凸集，$g_1(x),\cdots,g_m(x)$ 是 R^n 上的凸函数，$h_1(x),\cdots,h_p(x)$ 是 R^n 上的仿射函数，$\mathrm{ri}C\subset\mathrm{dom}g_i$，如果不存在 $x\in C$，满足

$$g_i(x)<0,\ i=1,\cdots,m,$$
$$h_j(x)=0,\ j=1,\cdots,p,$$

则存在不全为零的数 $\lambda_1,\cdots,\lambda_m,\mu_1,\cdots,\mu_p,\lambda_i\geqslant 0,i=1,\cdots,m$，满足

$$\sum_{i=1}^m \lambda_i g_i(x)+\sum_{j=1}^p \mu_j h_j(x)\geqslant 0,\ x\in C.$$

证明与定理 3.12 相似，略.

ii）最优性条件

现在研究凸规划

$$\text{(CP)}\quad\begin{aligned}&\min\ f(x)\\&\text{s.t}\ g_i(x)\leqslant 0,\quad i=1,\cdots,m, &(34)\\&\quad\ \ h_j(x)=0,\qquad j=1,\cdots,p. &(35)\\&\quad\ \ x\in S.\end{aligned}$$

其中 S 是 R^n 中的非空凸集，$f(x),g_i(x),i=1,\cdots,m$ 是 R^n 上的正常凸函数. $h_j(x),j=1,\cdots,p$ 是 R^n 上的仿射函数.

$X=\{x\mid g_i(x)\leqslant 0,i=1,\cdots,m,h_j(x)=0,j=1,\cdots,p,x\in S\}$ 是可行解集.

以后总设 $S\subset\mathrm{dom}g_i$，$\mathrm{ri}S\subset\mathrm{ri\ dom}g_i$，$S\subset\mathrm{dom}h_j$，$\mathrm{ri}S\subset\mathrm{ri\ dom}h_j$

定义 3.2 称

$$L(x,\lambda,\mu)=f(x)+\sum_{i=1}^m \lambda_i g_i(x)+\sum_{j=1}^p \mu_j h_j(x)\qquad(36)$$

是(CP)的 Lagrange 函数. 如果 $\lambda_i\geqslant 0,i=1,\cdots,m$，使 $L(x\lambda,\mu)$ 在 S 中的下确界有限且等于(CP)的极小值，则称 $\lambda_1,\cdots,\lambda_m,\mu_1,$

$\cdots,\mu_p$是(CP) 的 Kuhn-Tucker 乘子，或称 $\lambda=(\lambda_1,\cdots,\lambda_m)$，$\mu=(\mu_1,\cdots,\mu_p)$是 Kuhn-Tucker 向量.

从理论上看，Kuhn-Tucker 乘子的价值在于如果知道这样的乘子，就可以转换(CP)的运算. 这样，对于(CP) 问题不需确定(CP)的可行解并在其上求 $f(x)$极小值，而只要求出 $L(x,\lambda,\mu)$ 在 R^n 上的极小集，再从极小集中除去不满足约束的点就行了. 下面的定理正好说明了这一点.

定理 3.14 设 $\lambda_1,\cdots,\lambda_m,\mu_1,\cdots,\mu_p$是(CP)的 Kuhn-Tucker 乘子，$D$ 是 $L(x,\lambda,\mu)$在 R^n 中达到下确界的点集. $I=\{i|\lambda_i=0,1\leqslant i\leqslant m\}$. 又

$D_0=\{\bar{x}\in D|g_i(\bar{x})\leqslant 0,i\in I,g_i(\bar{x})=0,\ i\notin I,\ h_j(\bar{x})=0,\ j=1,\cdots,p\}$.

则 D_0是(CP)的全体极小值点的集合.

证明 因为 $\lambda_1,\cdots,\lambda_m,\mu_1,\cdots,\mu_p$是(CP) 的 Kuhn-Tucker 乘子，故 $\inf\limits_{x\in R^n}L(x,\lambda,\mu)=\min\limits_{x\in X}f(x)$. 对于 $\forall x\in X$，有

$$\lambda_i g_i(x)\leqslant 0,i=1,\cdots,m \tag{37}$$

$$\mu_j h_j(x)=0,j=1,\cdots,p \tag{38}$$

故 $f(x)+\lambda_1 g_1(x)+\cdots+\lambda_m g_m(x)+\mu_1 h_1(x)+\cdots+\mu_p h_p(x)\leqslant f(x)$，即

$$L(x,\lambda,\mu)\leqslant f(x). \tag{39}$$

当且仅当 $x\in X$，且 $\lambda_i g_i(x)=0,i=1,\cdots,m,\mu_j h_j(x)=0,j=1,\cdots,p$时(39)式的等号成立. 所以$f(x)$的极小集包含在 $L(x,\lambda,\mu)$的极小集之中.

容易证明，$f(x)$的极小集就是 D_0，但 $f(x)$的极小集是(CP)的全体极小值点的集合，所以 D_0是(CP) 的全体极小值点的集合. ▌

在一定的条件下，Kuhn-Tucker 乘子是存在的.

定理 3.15 设(CP)的极小值不是$-\infty$，存在$x_1 \in X$，$x_1 \in \mathrm{ri}\,S$，满足

$$g_i(x_1)<0, i=1,\cdots,m, \tag{40}$$

则(CP)的Kuhn Tucker乘子存在(不一定唯一).

证明 1) 不存在等式约束的情形. 设(CP)的极小值为α. 由已知条件，不等式组

$$g_i(x)<0, i=1,\cdots,m. \tag{41}$$

在$\mathrm{ri}S$中有解. 而不等式组

$$f(x)-\alpha<0, g_i(x)<0, i=1,\cdots,m. \tag{42}$$

在S中无解.故由定理3.10,存在不全为0的非负实数$\lambda_0,\lambda_1,\cdots,\lambda_m$,满足

$$\lambda_0(f(x)-\alpha)+\lambda_1 g_1(x)+\cdots+\lambda_m g_m(x)\geqslant 0, \quad \forall x \in S. \tag{43}$$

在(43)式中$\lambda_0 \neq 0$. 因为如果$\lambda_0=0$,则有

$$\lambda_1 g_1(x)+\cdots+\lambda_m g_m(x)\geqslant 0, \forall x \in S,$$

而$\lambda_1,\cdots,\lambda_m$不全为0，根据定理3.10,这与(41)有解矛盾. 不妨设$\lambda_0=1$, $L(x,\lambda)=f(x)+\lambda_1 g_1(x)+\cdots+\lambda_m g_m(x)$.

由(43)式,$L(x,\lambda)\geqslant\alpha, \forall x \in S$, 故$\inf L(x,\lambda)\geqslant\alpha$. 另一方面，$\forall x \in X$，$L(x,\lambda)\leqslant f(x)$，又有$\inf L(x,\lambda)\leqslant\alpha$. 所以，$\inf L(x,\lambda)=\alpha$. 由定义3.2,$\lambda_1,\cdots,\lambda_m$是Kuhn-Tucker乘子.

2) 存在等式约束的情形. 这时等式约束$h_j(x)=0$可以用等价的不等式约束

$$h_j(x)\leqslant 0, \quad -h_j(x)\leqslant 0, \tag{44}$$

代替,于是得到与(CP)等价的(CP_1):

$$\min f(x)$$

$$(CP_1) \quad \text{s.t} \quad g_i(x)\leqslant 0, i=1,\cdots,m, \tag{45}$$

$$h_j(x)\leqslant 0, j=1,\cdots,p, \tag{46}$$

$$-h_j(x)\leqslant 0, \tag{47}$$

$$x \in S.$$

应用 1)的推证(注意应用定理 3.10),(CP_1) 的 Kuhn-Tucker 乘子是使函数

$$f(x)+\sum_{i=1}^{m}\lambda_i g_i(x)+\sum_{j=1}^{p}\mu'_j h_j(x)$$

$$+\sum_{j=1}^{p}\mu''_j[-h_j(x)]$$

的下确界等于(CP_1)的极小值的非负实数 $\lambda_1,\cdots,\lambda_m,\mu'_1,\cdots,\mu'_p,\mu''_1,\cdots,\mu''_p$. 但(CP)的极小值与$(CP_1)$的极小值相同. 故如果设 $\mu_j=\mu'_j-\mu''_j$, $j=1,\cdots,p$,于是 $\lambda_1,\cdots,\lambda_m,\mu_1,\cdots,\mu_p$ 是(CP)的 Kuhn-Tucker 乘子. ▌

例 4 研究 R^2 中的(CP):

$$\begin{aligned}
&\min\ f(\xi_1,\xi_2)=\xi_1\\
&\text{s.t}\ \ g_1(\xi_1,\xi_2)=\xi_2\leqslant 0,\\
&\qquad g_2(\xi_1,\xi_2)=\xi_1^2-\xi_2\leqslant 0,\\
&\qquad S=R^2.
\end{aligned}$$

显然 $X=\{(0,0)\}$. 故 $x^*=(0,0)$是(CP)的唯一极小值点. 极小值是 0. 如果 λ_1,λ_2 是 Kuhn-Tucker 乘子,则

$$\begin{aligned}
&f(\xi_1,\xi_2)+\lambda_1 f_1(\xi_1,\xi_2)+\lambda_2 f_2(\xi_1,\xi_2)\\
&=\xi_1+(\lambda_1-\lambda_2)\xi_2+\lambda_2\xi_1^2\geqslant 0,\qquad \forall\ \xi_1,\xi_2
\end{aligned}$$

其中 $\lambda_1\geqslant 0$, $\lambda_2\geqslant 0$, 且 λ_1,λ_2 不全为 0. 这显然是不可能的. 故(CP)的 Kuhn-Tucker 乘子不存在. 可以验证, 在可行解集 X 中不存在满足

$$g_1(\xi_1,\xi_2)<0,\ g_2(\xi_1,\xi_2)<0,$$

的点 $x=(\xi_1,\xi_2)$,故定理 3.15 的条件不满足.

为了叙述的简洁, 在后面的三个定理中只考虑不等式约束的情形,这时 Lagrange 函数是 $L(x,\lambda)$.

定理 3.16 设(CP)的极小值不是$-\infty$，存在 $x_1\in X$，$x_1\in$ $\mathrm{ri}S$，满足

$$g_i(x_1)<0,\ i=1,\cdots,m. \tag{48}$$

则 $x^*\in X$ 是（CP）的极小值点的充分必要条件是存在非负实数 $\lambda_1,\cdots,\lambda_m$，满足

$$L(x^*,\lambda)\leqslant L(x,\lambda),\ x\in S. \tag{49}$$

$$\lambda_i g_i(x^*)=0,\ i=1,\cdots,m, \tag{50}$$

$$\lambda=(\lambda_1,\cdots,\lambda_m).$$

证明 设 x^* 是(CP)的极小值点. 根据定理 3.15，存在Kuhn-Tucker 乘子 $\lambda_1,\cdots\lambda_m$，这时当然有

$$f(x^*)\leqslant L(x,\lambda),\ x\in S. \tag{51}$$

因为 $x^*\in X$，则 $x^*\in S$，$g_i(x^*)\leqslant 0$. 所以

$$L(x^*,\lambda)=f(x^*)+\sum_{i=1}^{m}\lambda_i g_i(x^*)\leqslant f(x^*)\leqslant L(x^*,\lambda). \tag{52}$$

故

$$f(x^*)=L(x^*,\lambda), \tag{53}$$

$$\sum_{i=1}^{m}\lambda_i g_i(x^*)=0. \tag{54}$$

因为 $\lambda_i\geqslant 0$，$g_i(x^*)\leqslant 0$，则由(54)式得

$$\lambda_i g_i(x^*)=0,\ i=1,\cdots,m. \tag{55}$$

由(51)，(53)式知(49)式成立，

设 $x\in X$，且(49)，(50)式成立，则对 $\forall x\in X$，有

$$f(x)\geqslant f(x)+\sum_{i=1}^{m}\lambda_i g_i(x)=L(x,\lambda)$$

$$\geqslant L(x^*,\lambda)=f(x^*).$$

所以 x^* 是(CP)的极小值点. ∎

定理 3.17 设定理 3.16 的条件成立. $f(x)$, $g_i(x)$, $i=1$, $\cdots,m$,是 R^n 上的连续凸函数,则 x^* 是(CP) 的极小值点的充分必要条件是存在非负实数 $\lambda_1,\cdots,\lambda_m$, $\bar{x}\in K_S^*(x^*)$,满足

$$\bar{x}\in\partial f(x^*)+\sum_{i=1}^{m}\lambda_i\partial g_i(x^*), \tag{56}$$

$$\lambda_i g_i(x^*)=0,\ i=1,\cdots,m, \tag{57}$$

其中 $K_S(x^*)=\text{cone}(S-x^*)$.

证明 (49)式表明,x^* 是目标函数 $L(x,\lambda)$ 关于 x 在凸集 S 上的极小值点. 因为 $\lambda_i\geqslant 0$, $g_i(x)$ 是凸函数,故 $L(x,\lambda)$ 是凸连续函数. 由定理 3.5, $L(x,\lambda)$ 在 x^* 达到 S 上的极小值的充分必要条件是

$$\partial_x L(x^*,\lambda)\cap K_S^*(x^*)\neq\varnothing, \tag{58}$$

其中 $\partial_x L(x,\lambda)$ 是 $L(x,\lambda)$ 作为 x 的函数在 x 的次梯度. 而 $\partial_x L(x^*,\lambda)=\partial f(x^*)+\sum_{i=1}^{m}\lambda_i\partial g_i(x^*)$,故存在 $\bar{x}\in K_S^*(x^*)$, 满足

$$\bar{x}\in\partial f(x^*)+\sum_{i=1}^{m}\lambda_i\partial g_i(x^*). \tag{59}$$

故由定理 3.16, x^* 是(CP)的极小值点的充分必要条件是 (56), (57) 式成立. ▌

定理 3.18 设 $f(x)$, $g_i(x)$, $i=1,\cdots,m$,是 R^n 上的凸连续函数, x^* 是(CP)的极小值点,则存在不全为 0 的非负实数 $\lambda_i\geqslant 0$, $i=0,1,\cdots,m$,及 $\bar{x}\in\partial f(x^*)$, $\bar{x}_i\in\partial f_i(x^*)$,满足

$$\lambda_0\bar{x}+\sum_{i=1}^{m}\lambda_i\bar{x}_i\in K_S^*(x^*), \tag{60}$$

$$\lambda_i g_i(x^*)=0,\ i=1,\cdots,m, \tag{61}$$

$$\lambda_0 f(x^*)=0.$$

证明 设 $X_i=\{x\mid g_i(x)\leqslant 0\}, i=1,\cdots,m, X_0=\{x\mid f(x)\leqslant 0\}$. 则

$$X=\left(\bigcap_{i=1}^{m} X_i\right)\cap S.$$

设 x^* 是(CP)的极小值点,则由定理 3.7,存在

$$\bar{x}_i\in K^*_{X_i}(x^*),\ \ \bar{x}_S\in K^*_S(x^*),\ \ \bar{x}\in\partial f(x^*),$$

及 $\lambda_0\geqslant 0$,满足

$$\lambda_0\bar{x}=\sum_{i=1}^{m}\bar{x}_i+\bar{x}_S. \tag{62}$$

其中 $\lambda_0=0$ 时,$\bar{x}_i,\bar{x}_S$ 不同时为零向量.

因为 $x^*\in X_i$,这时有两种可能:

1) 存在 $x,g_i(x)<0$,由第三章定理 1.8 可知

$$K^*_{X_i}(x^*)=\begin{cases}\theta & g_i(x^*)<0,\\ -\text{cone}\,\partial g_i(x^*) & g_i(x^*)=0.\end{cases} \tag{63}$$

2) 不存在 x,使 $g_i(x)<0$,则由定理 3.3,$\theta\in\partial g_i(x^*)$.

于是可以分为两种情形讨论.

1) 对所有 i,在某个 $x,g_i(x)<0$,故(63)式成立. 这时设

$$\bar{x}_i=-\lambda_i\bar{x}_{i0},\ \ \lambda_i\geqslant 0,\ \ \bar{x}_{i0}\in\partial g_i(x^*), \tag{64}$$

$$\bar{x}_{00}=\bar{x},$$

同时满足

$$\lambda_i g_i(x^*)=0,\ \ i=1,\cdots,m, \tag{65}$$

即 $g_i(x^*)<0$ 时,$\lambda_i=0$;$g_i(x^*)=0$ 时,$\lambda_i\geqslant 0$. 因为 $x^*\in X_0$ 时,可按相似办法得 $\lambda_0 f(x^*)=0$,故将(64)式代入(62)式得

$$\lambda_0\bar{x}_{00}+\lambda_1\bar{x}_{10}+\cdots+\lambda_m\bar{x}_{m0}=\bar{x}_S, \tag{66}$$

$$\bar{x}_{i0}\in\partial g_i(x^*),\ \ \bar{x}_S\in K^*_S(x^*),\ \ \bar{x}_{00}\in\partial f(x^*).$$

$$\lambda_i\geqslant 0,\ \ \lambda_i g_i(x^*)=0,\ \ i=1,\cdots,m. \tag{67}$$

$$\lambda_0\geqslant 0,\ \ \lambda_0 f(x^*)=0. \tag{68}$$

因为 $\lambda_0=0$ 时，$\bar{x}_i,\bar{x}_S$ 不同时为零向量，所以 λ_i 不全为 0，故(60)、(61)式成立.

2) 设对某个 i_0，不存在 x，使 $g_{i0}(x)<0$，则 $\theta\in\partial f_{i0}(x^*)$. 这时设 $\lambda_{i0}=1,\bar{x}_{i0}=\theta$，而 $i\neq i_0$ 时 $\lambda_i=0,\bar{x}_S=\theta$. 这时(因为对 $f(x)$ 可相似确定 λ_0 及 $\bar{x}_{00}$)(66)-(68)式仍成立，故定理结论成立. ▌

iii) 可微假设下的最优性条件

ii) 中讨论的最优性条件，都没有假设目标函数及约束函数的可微性. 我们将会看到，在可微的假设下，本章§2所讨论过的某些必要条件也是充分的.

研究凸规划

$$\text{(CP)}\quad \begin{aligned} &\min\ f(x)\\ &\text{s.t}\quad g_i(x)\leqslant 0,\ i=1,\cdots,m, &(69)\\ &\qquad h_j(x)=0,\ j=1,\cdots,p, &(70)\\ &\qquad x\in S, \end{aligned}$$

其中 S 是 R^n 中的非空凸集，$f(x)$、$g_i(x)$，$i=1,\cdots,m$，是 R^n 上的可微凸函数，$h_j(x)$是 R^n 上的仿射函数，$j=1,\cdots,p$：

$$h_j(x)=\sum_{k=1}^{n}a_{jk}\xi_k-b_j,\quad x=(\xi_1,\cdots,\xi_n),\tag{71}$$

a_{jk},b_j 是实数.

定理 3.19 对于(69)、(70)式的(CP)，如果存在 $x^*\in X,\lambda_1,\cdots,\lambda_m,\mu_1,\cdots,\mu_p$，满足

$$\nabla_x L(x^*,\lambda,\mu)=\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)$$

$$+\sum_{j=1}^{p}\mu_j\nabla h_j(x^*)=\theta,\tag{72}$$

$$\lambda_i g_i(x^*)=0,\quad i=1,\cdots,m,\tag{73}$$

$$\lambda_i \geqslant 0,$$

则 x^* 是(CP)的极小值点.

证明 $\forall x \in X$,则

$$f(x) \geqslant f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x). \tag{74}$$

故由次梯度不等式，(74)式变成

$$f(x) \geqslant f(x^*) + \langle \nabla f(x^*), x - x^* \rangle + \sum_{i=1}^{m} \lambda_i g_i(x^*)$$

$$+ \sum_{i=1}^{m} \lambda_i \langle \nabla g_i(x^*), x - x^* \rangle + \sum_{j=1}^{p} \mu_j h_j(x^*)$$

$$+ \sum_{j=1}^{p} \mu_j \langle \nabla h_j(x^*), x - x^* \rangle.$$

重新调整右边的次序，有

$$f(x) \geqslant f(x^*) + \sum_{i=1}^{m} \lambda_i g_i(x^*) + \sum_{j=1}^{p} \mu_j h_j(x^*)$$

$$+ \langle \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*)$$

$$+ \sum_{j=1}^{p} \mu_j \nabla h_j(x^*), x - x^* \rangle.$$

考虑到(72),(73)式及 x^* 的可行性,得

$$f(x) \geqslant f(x^*).$$ ▌

定义 3.3 对于(CP),如果存在 $x \in X$,满足

$$g_i(x) < 0, \quad i = 1, \cdots, m, \tag{75}$$

$$h_j(x) = 0, \quad j = 1, \cdots, p, \tag{76}$$

则称(CP)是强相容的,或满足 Slater 条件.

定理 3.20 设向量组 $a_j = (a_{j1}, \cdots, a_{jn}), j = 1, \cdots, p$, 线

性无关，其中 a_{jk} 由(71)式决定．(CP) 强相容．如果 x^* 是(CP)的极小值点，则存在 $\lambda_1,\cdots,\lambda_m,\mu_1,\cdots,\mu_p$ 满足

$$\nabla_x L(x^*,\lambda,\mu)=\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)$$

$$+\sum_{j=1}^{p}\mu_j\nabla h_j(x^*)=\theta, \tag{77}$$

$$\lambda_i g_i(x^*)=0,\ i=1,\cdots,m, \tag{78}$$

$$\lambda_i\geqslant 0.$$

证明 因为 $f(x),g_i(x),i=1,\cdots,m,\ h_j(x),j=1,\cdots,p$ 的可微性，由第三章定理 3.6，知 $f(x)$，$g_i(x)$，$i=1,\cdots,m$，$h_j(x),j=1,\cdots,p$ 一阶连续可微．这满足定理 2.9′的条件，所以存在不全为 0 的实数 $\alpha_0,\alpha_1,\cdots,\alpha_m,\beta_1,\cdots,\beta_p$ 满足

$$\alpha_0\nabla f(x^*)+\sum_{i=1}^{m}\alpha_i\nabla g_i(x^*)+\sum_{j=1}^{p}\beta_j\nabla h_j(x^*)=\theta, \tag{79}$$

$$\alpha_i g_i(x^*)=0,\ i=1,\cdots,m, \tag{80}$$

$$\alpha_i\geqslant 0.$$

如果 $\alpha_0>0$，只要令 $\lambda_i=\alpha_i/\alpha_0,i=1,\cdots,m$，$\mu_j=\beta_j/\alpha_0,j=1,\cdots,p$，定理的结论自然成立．

假设 $\alpha_0=0$，这将导出矛盾．分两种情形．

1) $\alpha_0=0,\alpha_1,\cdots,\alpha_m$ 不全为 0．令 x_0 满足(75)，(76)式．则因为 $g_i(x)$ 是凸函数，故由(79)，(80)式，有

$$\sum_{i=1}^{m}\alpha_i g_i(x_0)\geqslant\sum_{i=1}^{m}\alpha_i g_i(x^*)$$

$$+\langle\sum_{i=1}^{m}\alpha_i\nabla g_i(x^*),x_0-x^*\rangle$$

$$\geq -\langle \sum_{j=1}^{p} \beta_j \nabla h_j(x^*), x_0 - x^* \rangle$$

$$\geq \sum_{j=1}^{p} \beta_j [h_j(x^*) - h_j(x_0)] = 0. \tag{81}$$

但 $g_i(x_0) < 0, i = 1, \cdots, m$, 而 $\alpha_i \geq 0, i = 1, \cdots, m$, 且 α_i 不全为0, 故 $\sum_{i=1}^{m} \alpha_i g_i(x_0) < 0$, 这与(81)式矛盾.

2) $\alpha_0 = 0, \alpha_i = 0, i = 1, \cdots, m$. 这时 $\beta_j, j = 1, \cdots, p$ 不全为0. 于是(79)式变成

$$\sum_{j=1}^{p} \beta_j \nabla h_j(x^*) = \theta,$$

或

$$\sum_{j=1}^{p} \beta_j a_{jk} = 0, \quad k = 1, \cdots, n. \tag{82}$$

(82)式表明向量组 $a_j = (a_{j1}, \cdots, a_{jn}), j = 1, \cdots, p$, 线性相关, 与已知条件$\{a_j\}$线性无关矛盾.

故 α_0 只能大于0. 如前所述, 定理的结论成立. ▌

§4 对偶问题与鞍点条件

作为一类约束极值问题, 存在另一类与之联系十分紧密的极值问题. 在一定的条件下, 两类问题具有相等的最优目标函数值, 而且可以通过求解一类问题而得到另一类问题的解答. 读者已在线性规划及其对偶规划中了解到这一点. 我们将这两类约束极值问题称为原始问题和对偶问题.

在最优化理论的发展过程中, 出现过不同形式的对偶问题, 这一节叙述 Lagrange 对偶问题的有关结果. 事实已经证明, 这种对

偶理论已在发展最优化算法等方面发挥了重大影响.

1. Lagrange 对偶问题

设已知(NLP):

$$
\text{(NLP)}\quad
\begin{aligned}
&\min\ f(x)\\
&\text{s.t}\ \ g_i(x)\leqslant 0,\ i=1,\cdots,m, &(1)\\
&\qquad h_j(x)=0,\ j=1,\cdots,p, &(2)\\
&\qquad x\in S.
\end{aligned}
$$

令 $\Phi(\lambda,\mu)=\inf\limits_{x\in S}\{f(x)+\sum\limits_{i=1}^{m}\lambda_i g_i(x)+\sum\limits_{j=1}^{p}\mu_j h_j(x)\}$,

其中 $\lambda=(\lambda_1,\cdots,\lambda_m)$, $\mu=(\mu_1,\cdots,\mu_p)$, 仍以 X 表示可行解集.

定义 4.1 称

$$
\text{(DNLP)}\quad
\begin{aligned}
&\max\ \Phi(\lambda,\mu)\\
&\text{s.t}\ \ \lambda\geqslant 0.
\end{aligned}
$$

是(NLP)的 Lagrange 对偶问题(DNLP). 而(NLP) 称为原始问题.

例 1 研究 R^2 中的(NLP):

$$
\begin{aligned}
&\min\ f(\xi_1,\xi_2)=\xi_1^2+\xi_2^2\\
&\text{s.t}\ \ -\xi_1-\xi_2+4\leqslant 0,\\
&\qquad -\xi_1\leqslant 0,\\
&\qquad -\xi_2\leqslant 0.
\end{aligned}
$$

其极小值点 $x^*=(2,2)$, $f(x^*)=8$.

令 $g(\xi_1,\xi_2)=-\xi_1-\xi_2+4$, $S=\{(\xi_1,\xi_2)\,|\,\xi_1\geqslant 0,\xi_2\geqslant 0\}$.

则

$$
\begin{aligned}
\Phi(\lambda)&=\inf\{\xi_1^2+\xi_2^2+\lambda(-\xi_1-\xi_2+4)\,|\,\xi_1\geqslant 0,\xi_2\geqslant 0\}\\
&=\inf\{\xi_1^2-\lambda\xi_1\,|\,\xi_1\geqslant 0\}+\inf\{\xi_2^2-\lambda\xi_2+4\lambda\,|\,\xi_2\geqslant 0\}
\end{aligned}
$$

$$=\begin{cases}-\frac{1}{2}\lambda^2+4\lambda & \lambda\geqslant 0,\\ 4\lambda & \lambda<0.\end{cases}$$

$\Phi(\lambda)$是凹函数，$\lambda\geqslant 0$ 时极大值在 $\lambda^*=4$ 取得，$\Phi(\lambda^*)=8$. 这是(DNLP)的结果.

从例 1 已经看出，原始问题与对偶问题的最优目标函数值是相等的. 在一般情况下，需要一定的条件才能保证做到这一点.

定理 4.1 （弱对偶定理）设 $\bar{x}$ 是(NLP)的可行解，λ,μ 是(DNLP)的可行解，则

$$f(\bar{x})\geqslant\Phi(\lambda,\mu). \tag{3}$$

证明 因为 $\bar{x}\in X$，故 $g_i(\bar{x})\leqslant 0, i=1,\cdots,m$，$h_j(\bar{x})=0$，$j=1,\cdots,p$，且 $\lambda=(\lambda_1,\cdots,\lambda_m)\geqslant 0$. 于是由 $\Phi(\lambda,\mu)$ 的定义，有

$$\Phi(\lambda,\mu)=\inf\left\{f(x)+\sum_{i=1}^{m}\lambda_i g_i(x)+\sum_{j=1}^{p}\mu_j h_j(x)\,\middle|\,x\in S\right\}$$

$$\leqslant f(\bar{x})+\sum_{i=1}^{m}\lambda_i g_i(\bar{x})+\sum_{j=1}^{p}\mu_j h_j(\bar{x})$$

$$\leqslant f(\bar{x}),$$

即(3)成立. ▮

推论 4.1.1 在(NLP)和(DNLP)中，有

$$\inf\{f(x)\,|\,x\in X\}\geqslant\sup\{\Phi(\lambda,\mu)\,|\,\lambda\geqslant 0\}. \tag{4}$$

证明留给读者.

推论 4.1.2 如果在(NLP)及(DNLP)中，$f(\bar{x})\leqslant\Phi(\bar{\lambda},\bar{\mu})$，其中 $\bar{\lambda}\geqslant 0$，$\bar{x}\in X$. 则 $\bar{x}$ 是(NLP)的极小值点，$\bar{\lambda},\bar{\mu}$ 是(DNLP)的极大值点.

证明留给读者.

推论 4.1.3 在(NLP)及(DNLP)中，如果 $\inf\limits_{x\in X} f(x)=$

$-\infty$,则对 $\forall\lambda\geqslant 0,\Phi(\lambda,\mu)=-\infty$;如果 $\sup\limits_{\lambda\geqslant 0}\Phi(\lambda,\mu)=+\infty$,则(NLP)没有可行解($X=\varnothing$).

证明留给读者.

例 2 研究 R^2的(NLP):

$$\begin{aligned}
\min\ & f(\xi_1,\xi_2)=-2\,\xi_1+\xi_2\\
\text{s.t}\ & \xi_1+\xi_2-3=0,\\
& (\xi_1,\xi_2)\in S.
\end{aligned}$$

其中. $S=\{(0,0),(0,4),(4,4),(4,0),(1,2),(2,1)\}$.
容易验证 $(\xi_1^*,\ \xi_2^*)=(2,1)$ 是原始问题的极小值点,$f(\xi_1^*,\ \xi_2^*)=-3$.

$$\Phi(\mu)=\min\{-2\,\xi_1+\xi_2+\mu(\xi_1+\xi_2-3)\,|\,(\xi_1,\xi_2)\in S\}$$

$$=\begin{cases}-4+5\,\mu & \mu\leqslant -1,\\ -8+\mu & -1\leqslant\mu\leqslant 2,\\ -3\,\mu & \mu\geqslant 2.\end{cases}$$

对偶问题的极大值点是 $\mu^*=2$,$\Phi(\mu^*)=-6$.

这个例子说明,(NLP)与(DNLP)的最优目标函数值不一定相等.

定义 4.2 如果原始问题的最优目标函数值大于对偶问题的最优目标函数值,则称存在对偶间隙.

由以上定义,例 2 存在对偶间隙.

下面的定理表明,在目标函数及约束函数的凸性条件下,对偶间隙是可以消除的.

定理 4.2(强对偶定理) 设 S 是 R^n 中的非空凸集,$f(x)$,$g_i(x),i=1,\cdots,m,h_j(x),j=1,\cdots,p$ 分别是 R^n 上的凸函数和仿射函数.如果Slater 条件成立,即存在 $\bar{x}\in X,g_i(\bar{x})<0,i=1,\cdots,m,h_j(\bar{x})=0,j=1,\cdots,p$,且 $\theta\in\operatorname{int}C$,$C=\{(h_1(x),\cdots,h_p(x))\,|\,x\in X\}$,则

$$\inf_{x\in X} f(x) = \sup_{\lambda\geq 0,\mu}\Phi(\lambda,\mu). \tag{5}$$

如果 $\inf\limits_{x\in X} f(x)$有限，则 $\sup\limits_{\lambda\geq 0,\mu}\Phi(\lambda,\mu)$在$(\lambda^*,\mu^*)$达到，$\lambda^*\geq\theta$；

如果 $\inf\limits_{x\in X} f(x)$在 x^* 达到，则$\sum\limits_{i=1}^{m}\lambda_i^* g_i(x^*)=0$。$\lambda^*=(\lambda_1^*,\cdots,\lambda_m^*)$，$\mu^*=(\mu_1^*,\cdots,\mu_p^*)$。

证明 令 $\inf\limits_{x\in X} f(x)=\alpha$。

如果 $\alpha=-\infty$，则由推论 4.1.3，$\sup\limits_{\lambda\geq 0,\mu}\Phi(\lambda,\mu)=-\infty$。故(5) 式成立.

设α 有限．这时不等式及方程组

$$f(x)-\alpha<0, g_i(x)\leq 0, i=1,\cdots,m,$$
$$h_j(x)=0, j=1,\cdots,p, x\in X.$$

无解．根据定理 3.13，存在不全为 0 实数 $\alpha_i\geq 0, i=0,1,\cdots,m$，$\beta_j, j=1,\cdots,p$，对于$\forall x\in X$，有

$$\alpha_0(f(x)-\alpha)+\sum_{i=1}^{m}\alpha_i g_i(x)+\sum_{j=1}^{p}\beta_j h_j(x)\geq 0. \tag{6}$$

假设 $\alpha_0=0$，由 Slater 条件，(6)式变为

$$\sum_{i=1}^{m}\alpha_i g_i(x)\geq 0.$$

但应有$\sum\limits_{i=1}^{m}\alpha_i g_i(x)\leq 0$，故 $\alpha_i=0, i=1,\cdots,m$。故(6)式为

$$\sum_{j=1}^{p}\beta_j h_j(x)\geq 0, \forall x\in X. \tag{7}$$

因为 $\theta\in\operatorname{int}C$，故可以取一个 $x\in X$，使$(h_1(x),\cdots,h_p(x))=-k(\beta_1,\cdots,\beta_p)$，$k>0$．对这个 x，(7)式成为

$$\sum_{j=1}^{p}\beta_j h_j(x)=-k\sum_{j=1}^{p}\beta_j^2\geqslant 0, \tag{8}$$

故 $\beta_j=0, j=1,\cdots,p$. 这与$\alpha_i, i=0,1,\cdots,m, \beta_j, j=1,\cdots,p$, 不全为 0 矛盾. 所以 $\alpha_0>0$.

令 $\lambda_i^*=\alpha_i/\alpha_0, i=1,\cdots,m$. $\mu_j^*=\beta_j/\alpha_0, j=1,\cdots,p$. 则(6) 式变成

$$f(x)+\sum_{i=1}^{m}\lambda_i^* g_i(x)+\sum_{j=1}^{p}\mu_j^* h_j(x)\geqslant\alpha, \forall x\in X. \tag{9}$$

于是

$$\Phi(\lambda^*,\mu^*)=\inf\{f(x)+\sum_{i=1}^{m}\lambda_i^* g_i(x)$$

$$+\sum_{j=1}^{p}\mu_j^* h_j(x)\,|\,x\in X\}\geqslant\alpha.$$

由推论 4.1.2,得 $\Phi(\lambda^*,\mu^*)=\alpha$, 故$(\lambda^*,\mu^*)$ 为对偶问题的最优解.

最后,设 x^*是原始问题的解, 即 $g_i(x^*)\leqslant 0$, $i=1,\cdots,m$, $h_j(x^*)=0, j=1,\cdots,p, f(x^*)=\alpha$. 在(9)中令 $x=x^*$,则有

$$\sum_{i=1}^{m}\lambda_i^* g_i(x^*)\geqslant 0.$$

而 $\sum_{i=1}^{m}\lambda_i^* g_i(x^*)\leqslant 0$, 故 $\sum_{i=1}^{m}\lambda_i^* g_i(x^*)=0$. ▌

2. 鞍点问题

定义 4.3 设 $\psi(x,y)$是 $x\in R^n, y\in R^m$ 的实函数,点$(\bar{x},\bar{y})$满足:对 $\forall x\in R^n,\forall y\in R^m$,有

$$\psi(\bar{x},y)\leqslant\psi(\bar{x},\bar{y})\leqslant\psi(x,\bar{y}), \tag{10}$$

则称$(\bar{x},\bar{y})$是 $\psi(x,y)$的鞍点.

定义 4.4 对于(NLP)，如果$\bar{x}\in R^n,\bar{\lambda}\in R^m,\bar{\mu}\in R^p,\bar{\lambda}\geqslant\theta$，使 Lagrange 函数 $L(x,\lambda,\mu)=f(x)-\sum\limits_{i=1}^{m}\lambda_i g_i(x)-\sum\limits_{j=1}^{p}\mu_j h_j(x)$ 满足：对 $\forall x\in R^n,\forall\lambda\in R^m,\forall\mu\in R^p,\lambda\geqslant\theta$，有

$$L(\bar{x},\lambda,\mu)\leqslant L(\bar{x},\bar{\lambda},\bar{\mu})\leqslant L(x,\bar{\lambda},\bar{\mu}), \tag{11}$$

则称$(\bar{x},\bar{\lambda},\bar{\mu})$是 $L(x,\lambda,\mu)$的鞍点，求 $L(x,\lambda,\mu)$ 的鞍点问题称为(SP)，其中 $\lambda=(\lambda_1,\cdots,\lambda_m),\bar{\lambda}=(\bar{\lambda}_1,\cdots,\lambda_m),\mu=(\mu_1,\cdots,\mu_p),\bar{\mu}=(\bar{\mu}_1,\cdots,\bar{\mu}_p)$.

定理 4.3 如果$(\bar{x},\bar{\lambda},\bar{\mu})$是(SP)的解，则 $\bar{x}$ 和 $\bar{\lambda}$，$\bar{\mu}$ 分别是(NLP)和(DNLP)的解.

证明 设$(\bar{x},\bar{\lambda},\bar{\mu})$是(SP)的解，由(11)式知

$$\begin{aligned}f(\bar{x})-\sum_{i=1}^{m}\lambda_i g_i(\bar{x})-\sum_{j=1}^{p}\mu_j h_j(\bar{x})&\leqslant f(\bar{x})\\+\sum_{i=1}^{m}\bar{\lambda}_i g_i(\bar{x})-\sum_{j=1}^{p}\bar{\mu}_j h_j(\bar{x})&\leqslant f(x)-\sum_{j=1}^{m}\bar{\lambda}_i g_i(x)\\+\sum_{j=1}^{p}\bar{\mu}_j h_j(x).\end{aligned} \tag{12}$$

由(12)式左边的不等式，对于 $\forall\lambda\in R^m,\lambda\geqslant\theta,\forall\mu\in R^p$，有

$$\sum_{i=1}^{m}(\lambda_i-\bar{\lambda}_i)g_i(\bar{x})+\sum_{j=1}^{p}(\mu_j-\bar{\mu}_j)h_j(\bar{x})\leqslant 0. \tag{13}$$

下面证明 $h_j(\bar{x})=0,j=1,\cdots,p$.

设对于满足 $1\leqslant k\leqslant p$ 的一个 $k,h_k(\bar{x})>0$，则在(13)式中左边令 $\lambda_i=\bar{\lambda}_i,i=1,\cdots,m,\mu_j=\bar{\mu}_j,j\neq k,\mu_k=\bar{\mu}_k+1$，得到与(13)式矛盾的结果：

$$\sum_{i=1}^{m}(\lambda_i-\bar{\lambda}_i)g_i(\bar{x})+\sum_{j=1}^{p}(\mu_j-\bar{\mu}_j)h_j(\bar{x})>0.$$

如果 $h_k(\bar{x})<0$,同样有矛盾的结果,故 $h_j(\bar{x})=0, j=1, \cdots, p$.

设 $\mu_j=\bar{\mu}_j, \lambda_1=\bar{\lambda}_1+1, \lambda_i=\bar{\lambda}_i, i=2, \cdots, m$. 从(13)式又得 $g_1(\bar{x})\leqslant 0$. 重复这个过程,得 $g_i(\bar{x})\leqslant 0$. $i=2,\cdots,m$. 所以 $\bar{x}\in X$.

在(13)式中如果令 $\lambda=\theta$. 又得 $-\sum_{i=1}^{m}\bar{\lambda}_i g_i(\bar{x})\leqslant 0$. 但 $\bar{\lambda}_i\geqslant\theta, g_i(\bar{x})\leqslant 0, i=1,\cdots,m$, 所以 $\sum_{i=1}^{m}\bar{\lambda}_i g_i(\bar{x})=0$. 因而

$$\bar{\lambda}_i g_i(\bar{x})=0, i=1,\cdots,m. \tag{14}$$

由(12)式右边的不等式及上面所证的结果，有

$$f(\bar{x})\leqslant f(x)+\sum_{i=1}^{m}\bar{\lambda}_i g_i(x)+\sum_{j=1}^{p}\bar{\mu}_j h_j(x). \tag{15}$$

如果 $x\in X$,即 $g_i(x)\leqslant 0, h_j(x)=0, i=1,\cdots,m, j=1,\cdots,p$, 则

$$f(\bar{x})\leqslant f(x).$$

故 $\bar{x}$ 是(NLP)的极小值点. 而从(10)式又表示

$$f(\bar{x})\leqslant\Phi(\bar{\lambda},\bar{\mu}).$$

而 $\bar{\lambda}\geqslant\theta$,根据推论 4.1.2,$\bar{\lambda},\bar{\mu}$ 是(DNLP)的最优解. ▌

定理 4.4 设 S 是 R^n 中的非空凸集. $f(x), g_i(x), i=1,\cdots,m$ 是 R^n 上的凸函数,$h_j(x)$ 是 R^n 上的仿射函数,$j=1,\cdots,p$. 又设 Slater 条件成立,$\theta\in\operatorname{int}C, C=\{(h_1(x),\cdots,h_p(x))|x\in X\}$. 如果 $\bar{x}$ 是(NLP)的极小值点, 则存在 $\bar{\lambda}=(\bar{\lambda}_1,\cdots,\bar{\lambda}_m)\geqslant\theta, \bar{\mu}=(\bar{\mu}_1,\cdots,\bar{\mu}_p)$, 使$(\bar{x},\bar{\lambda},\bar{\mu})$是(SP)的解. 即

$$L(\bar{x},\lambda,\mu)\leqslant L(\bar{x},\bar{\lambda},\bar{\mu})\leqslant L(x,\bar{\lambda},\bar{\mu}), \tag{16}$$

且

$$\bar{\lambda}_i g_i(\bar{x})=0, i=1,\cdots,m. \tag{17}$$

证明 由定理 4.2 证明的前半部分知存在 $\bar{\lambda}=(\bar{\lambda}_1,\cdots,\bar{\lambda}_m)$

$\geqslant\theta, \bar{\mu}=(\bar{\mu}_1,\cdots,\bar{\mu}_p)$，对于 $\forall x\in X$，有

$$f(\bar{x})\leqslant f(x)+\sum_{i=1}^{m}\bar{\lambda}_i g_i(x)+\sum_{j=1}^{p}\bar{\mu}_j h_j(x)$$

$$=L(x,\bar{\lambda},\bar{\mu}). \tag{18}$$

同样由定理 4.2，$\sum_{i=1}^{m}\bar{\lambda}_i g_i(\bar{x})=0$，而 $h_j(\bar{x})=0,\ j=1,\cdots,p$，故

$$f(\bar{x})=f(\bar{x})+\sum_{i=1}^{m}\bar{\lambda}_i g_i(\bar{x})+\sum_{j=1}^{p}\bar{\mu}_j h_j(\bar{x})=L(\bar{x},\bar{\lambda},\bar{\mu}). \tag{19}$$

但 $g_i(\bar{x})\leqslant 0, h_j(\bar{x})=0, i=1,\cdots,m, j=1,\cdots,p$，又有

$$L(\bar{x},\lambda,\mu)=f(\bar{x})+\sum_{i=1}^{m}\lambda_i g_i(\bar{x})+\sum_{j=1}^{p}\mu_j h_j(\bar{x})$$

$$\leqslant f(\bar{x}),\forall\lambda_i\geqslant 0,i=1,\cdots,m. \tag{20}$$

结合(18)-(20)式得(16)式，(17)式成立. ▌

下面两个定理讨论了鞍点条件与§2的 Kuhn-Tucker 条件的关系.

定理 4.5 在(NLP)中，设 $f(x), g_i(x), i=1,\cdots,m$，是 R^n 上的可微凸函数，$h_j(x), j=1,\cdots,p$，是 R^n 上的仿射函数. 如果 $\bar{x}\in X$ 满足 Kuhn-Tucker 条件，即存在 $\bar{\lambda}=(\bar{\lambda}_1,\cdots,\bar{\lambda}_m)\geqslant\theta$，$\bar{\mu}=(\bar{\mu}_1,\cdots,\bar{\mu}_p)$，满足

$$\nabla f(\bar{x})+\sum_{i=1}^{m}\bar{\lambda}_i\nabla g_i(\bar{x})+\sum_{j=1}^{p}\bar{\mu}_j\nabla h_j(\bar{x})=\theta, \tag{21}$$

$$\bar{\lambda}_i g_i(\bar{x})=0, i=1,\cdots,m. \tag{22}$$

则 $(\bar{x},\bar{\lambda},\bar{\mu})$ 是(SP)的解.

证明　由 $f(x), g_i(x), i=1,\cdots,m, h_j(x), j=1,\cdots,p$ 的

凸性及可微性知.对 $\forall x\in X$,有

$$f(x)\geqslant f(\bar{x})+\langle\nabla f(\bar{x}),x-\bar{x}\rangle,\tag{23}$$

$$g_i(x)\geqslant g_i(\bar{x})+\langle\nabla g_i(\bar{x}),x-\bar{x}\rangle,\tag{24}$$

$$h_j(x)=h_j(\bar{x})+\langle\nabla h_j(\bar{x}),x-\bar{x}\rangle.\tag{25}$$

用 $\bar{\lambda}_i$ 乘(24),$\bar{\mu}_j$ 乘(25),再与(23)相加,得

$$f(x)+\sum_{i=1}^{m}\bar{\lambda}_i g_i(x)+\sum_{j=1}^{p}\bar{\mu}_j h_j(x)$$

$$\geqslant f(\bar{x})+\sum_{i=1}^{m}\bar{\lambda}_i g_i(\bar{x})+\sum_{j=1}^{p}\bar{\mu}_j h_j(\bar{x})+$$

$$\langle\nabla f(\bar{x})+\sum_{i=1}^{m}\bar{\lambda}_i\nabla g_i(\bar{x})+\sum_{j=1}^{p}\bar{\mu}_j\nabla h_j(\bar{x}),x-\bar{x}\rangle$$

$$=f(\bar{x})+\sum_{i=1}^{m}\bar{\lambda}_i g_i(\bar{x})+\sum_{j=1}^{p}\bar{\mu}_j h_j(\bar{x}),$$

即

$$L(\bar{x},\bar{\lambda},\bar{\mu})\leqslant L(x,\bar{\lambda},\bar{\mu}).\tag{26}$$

但 $g_i(\bar{x})\leqslant 0,i=1,\cdots,m,h_j(\bar{x})=0,j=1,\cdots,p$,又有

$$L(\bar{x},\lambda,\mu)\leqslant L(\bar{x},\bar{\lambda},\bar{\mu}),\forall\lambda=(\lambda_1,\cdots,\lambda_m)\geqslant\theta.\tag{27}$$

结合(26),(27)式知$(\bar{x},\bar{\lambda},\bar{\mu})$是(SP)的解. ▌

定理 4.6 在 (NLP) 中, 设 $f(x)$, $g_i(x)$, $i=1,\cdots,m$, $h_j(x)$, $j=1,\cdots,p$是R^n上的可微实函数. 如果$(\bar{x},\bar{\lambda},\bar{\mu})$是(SP)的解,则$E(\bar{x})\cap F(\bar{x})=\varnothing$,且

$$\nabla_x L(\bar{x},\bar{\lambda},\bar{\mu})=\nabla f(\bar{x})+\sum_{i=1}^{m}\bar{\lambda}_i\nabla g_i(\bar{x})$$

$$+\sum_{j=1}^{p}\bar{\mu}_j\nabla h_j(\bar{x})=\theta,\tag{28}$$

$$\bar{\lambda}_i g_i(\bar{x})=0,\bar{\lambda}_i\geqslant 0,i=1,\cdots,m,\tag{29}$$

其中 $E(x)$, $F(x)$的定义见定义 2.4.

证明 在定理 4.3 已经证明, 如果$(\bar{x},\bar{\lambda},\bar{\mu})$是(SP) 的解, 则 $\bar{\lambda}_i g_i(\bar{x})=0, i=1,\cdots,m$, 即(29)成立. 由 (SP) 的解的定义, 自然有 $\bar{\lambda}_i \geqslant 0, i=1,\cdots,m$.

由定义 4.4 中 (11) 式右边的不等式, 对每一个 $y\in R^n$, 当 $\alpha>0$ 充分小时, 有

$$L(\bar{x},\bar{\lambda},\bar{\mu})\leqslant L(\bar{x}+\alpha y,\bar{\lambda},\bar{\mu}).$$

故

$$\frac{L(\bar{x}+\alpha y,\bar{\lambda},\bar{\mu})-L(\bar{x},\bar{\lambda},\bar{\mu})}{\alpha}\geqslant 0. \tag{30}$$

在 $\alpha\to 0$ 时取极限, 得

$$\langle \nabla_x L(\bar{x},\bar{\lambda},\bar{\mu}), y\rangle \geqslant 0. \tag{31}$$

考虑到 y 在 R^n 的任意性, 得(28)式由定理 2.6, 得

$$E(\bar{x})\cap F(\bar{x})=\varnothing. \ \blacksquare$$

从上面的两个定理表明, 在凸性的条件下, 如果 $\bar{x}$ 是 Kuhn-Tucker 点, 则 Kuhn-Tucker 条件中的 Lagrange 乘子也是鞍点条件中的乘子. 反之, 鞍点条件中的乘子是 Kuhn-Tucker 条件中的 Lagrange 乘子.

3. $\psi(x,y)$的极小—极大问题.

定义 4.5 设 $\psi(x,y)$如定义 4.3, 求函数 $\psi(x,y)$ 关于 x 的极小值及关于 y 的极大值的问题称为 $\psi(x,y)$ 的极小—极大问题.

定理 4.7 设 $\psi(x,y)$ 是两个变向量 $x\in D\subset R^n$ 和 $y\in E\subset R^m$ 的实函数, 则

$$\max_{y\in E}\min_{x\in D}\psi(x,y)\leqslant\min_{x\in D}\max_{y\in E}\psi(x,y), \tag{32}$$

只要(32)式的极小和极大存在.

证明 对于任意固定的 $y\in E$, 有

$$\min_{x\in D}\psi(x,y)\leqslant\psi(x,y). \tag{33}$$

相似地，对任意固定的 $x\in D$，有

$$\psi(x,y)\leqslant\max_{y\in E}\psi(x,y). \tag{34}$$

故有

$$\min_{x\in D}\psi(x,y)\leqslant\max_{y\in E}\psi(x,y). \tag{35}$$

进而有

$$\max_{y\in E}\min_{x\in D}\psi(x,y)\leqslant\min_{x\in D}\max_{y\in E}\psi(x,y).$$

下面的定理说明了极小一极大问题与鞍点的关系.

定理 4.8 设定理 4.7 的条件成立，则

$$\max_{y\in E}\min_{x\in D}\psi(x,y)=\psi(\bar{x},\bar{y})=\min_{x\in D}\max_{y\in E}\psi(x,y) \tag{36}$$

的等价条件是$(\bar{x},\bar{y})$是 $\psi(x,y)$的鞍点.

证明 设$(\bar{x},\bar{y})$是 $\psi(x,y)$的鞍点，则由定义 4.3，有

$$\max_{y\in E}\psi(\bar{x},y)\leqslant\psi(\bar{x},\bar{y})\leqslant\min_{x\in D}\psi(x,\bar{y}). \tag{37}$$

而

$$\min_{x\in D}\max_{y\in E}\psi(x,y)\leqslant\max_{y\in E}\psi(\bar{x},y), \tag{38}$$

$$\min_{x\in D}\psi(x,\bar{y})\leqslant\max_{y\in E}\min_{x\in D}\psi(x,y). \tag{39}$$

由(37),(38),(39)式有

$$\min_{x\in D}\max_{y\in E}\psi(x,y)\leqslant\psi(\bar{x},\bar{y})\leqslant\max_{y\in E}\min_{x\in D}\psi(x,y). \tag{40}$$

与(32)式比较，知(36)式成立.

反之，设$(\bar{x},\bar{y})$满足

$$\max_{y\in E}\psi(\bar{x},y)=\min_{x\in D}\max_{y\in E}\psi(x,y), \tag{41}$$

$$\min_{x\in D}\psi(x,\bar{y})=\max_{y\in E}\min_{x\in D}\psi(x,y). \tag{42}$$

如果(36)成立，则

$$\begin{aligned}\psi(\bar{x},y)&\leqslant\max_{y\in E}\psi(\bar{x},y)=\psi(\bar{x},\bar{y})\\&=\min_{x\in D}\psi(x,\bar{y})\leqslant\psi(x,\bar{y}).\end{aligned}$$

故$(\bar{x},\bar{y})$是$\psi(x,y)$的鞍点. ▌

习　　题

1. 证明推论 1.2.1.
2. 证明定理 1.4.
3. 证明定理 1.5.
4. 证明定理 1.7.
5. 证明推论 1.9.1.
6. 证明定理 1.10.
7. 证明定理 1.11.
8. 证明推论 1.12.1.
9. 证明定理 2.4.
10. 研究无约束极值问题:

$$\min_{x\in R^2} f(x)=\xi_1^2-\xi_1\xi_2+2\xi_2^2-2\xi_1+e^{\xi_1+\xi_2},\quad x=(\xi_1,\xi_2).$$

1) 写出一阶必要条件. 这个条件对最优性是否充分? 为什么?

2) $\bar{x}=(0,0)$是否为极小值点? 如果不是,找出函数减小的一个方向y.

3) 由$\bar{x}=(0,0)$出发,沿2)所找出的方向y求目标函数的极小值.

11. 研究约束极值问题:

$$\begin{aligned}\min\quad & f(x)\\ \text{s.t}\quad & \xi_1^2+(\xi_2-1)^2\leqslant 1,\\ & \xi_2-\xi_1\leqslant 0,\ x=(\xi_1,\xi_2).\end{aligned}$$

设$\bar{x}=(0,0)$. 求一个$f(x)$,使在$\bar{x}$,有$E(\bar{x})\cap F(\bar{x})=\varnothing$,但$\bar{x}$不是上述问题的极小值点.

12. 研究(NLP):

$$\begin{aligned}\min\ & f(x)=(\xi_1-3)^2+(\xi_2-4)^2\\ \text{s.t}\quad & \xi_1^2+\xi_2^2-8\leqslant 0,\\ & 2\xi_1-\xi_2^2\geqslant 0,\ x=(\xi_1,\xi_2).\end{aligned}$$

试检验$x_1=(2,2)$,$x_2=(2,-2)$是否满足定理2.7的结论.

13. 研究(NLP):

$$\max\ f(x)=4\xi_1+2\xi_2$$

$$\text{s.t}\quad \xi_1\log\xi_1-\xi_2+1\geqslant 0,$$
$$\xi_1\geqslant 0,$$
$$3\xi_1-2\xi_2^2\geqslant 1,\ x=(\xi_1,\xi_2).$$

1）证明不存在满足 Kuhn-Tucker 必要条件的点 $\bar{x}\in R^2$。

2）存在满足 Fritz Zohn 必要条件的点吗？

14．研究(NLP)：

$$\min f(x)=\left(\xi_1-\frac{9}{2}\right)^2+(\xi_2-2)^2$$
$$\text{s.t}\quad \xi_2-\xi_1^2\geqslant 0,$$
$$\xi_1+\xi_2\leqslant 6,$$
$$\xi_1\geqslant 0,$$
$$\xi_2\geqslant 0,\ x=(\xi_1,\xi_2).$$

1）写出它的 Kuhn-Tucker 必要条件，并验证在点 $\bar{x}=\left(\frac{3}{2},\ \frac{9}{2}\right)$这个条件成立.

2）证明 $\bar{x}$ 是唯一的全局极小值点。

15．研究(NLP)：

$$\min\quad f(x)=-3\xi_1+\xi_2-\xi_3^2$$
$$\text{s.t}\quad \xi_1+\xi_2+\xi_3\leqslant 0,$$
$$-\xi_1+2\xi_2+\xi_3^2=0,\ x=(\xi_1,\xi_2,\xi_3).$$

1）写出它的 Kuhn-Tucker 必要条件。

2）用上述条件求出局部极小值点。

16．研究线性规划问题：

$$\max\quad f(x)=2\xi_1+3\xi_2$$
$$\text{s.t}\quad \xi_1+\xi_2\leqslant 8,$$
$$-\xi_1+2\xi_2\leqslant 4,$$
$$\xi_1\geqslant 0,$$
$$\xi_2\geqslant 0,\ x=(\xi_1,\xi_2).$$

1）写出它的 Kuhn-Tucker 必要条件.

2）对可行集的每个极点，验证上述条件是否成立，并由此求出极大值点.

17. 研究(NLP)：

$$\min f(x)$$
$$\text{s.t} \quad x \geqslant 0, x \in R^n.$$

证明上述问题的 Kuhn-Tucker 必要条件是：

$$\langle \nabla f(x^*), x^* \rangle = 0, \nabla f(x^*) \geqslant \theta,$$

其中 x^* 是局部极小值点.

18. 研究(NLP)：

$$\max f(x)$$
$$\text{s.t} \quad g_i(x) \geqslant 0, i=1,\cdots,k.$$
$$g_i(x) \leqslant 0, i=k+1,\cdots,m$$
$$h_j(x)=0, \ j=1,\cdots,p.$$

试叙述它的 Kuhn-Tucker 必要条件.

19. 研究一维极小化问题：

$$\min f(x+\lambda d)$$
$$\text{s.t} \quad \lambda \geqslant 0,$$

其中 x 是 R^n 中的固定点，$d \in R^n$ 是已知的非零方向.

1) 设 $f(x)$ 可微，写出它的最优性必要条件. 这个条件充分吗？如果不是，对 $f(x)$ 作什么假定可以使必要条件又是充分条件？

2) 如果 $f(x)$ 是凸函数但不一定可微，试利用次梯度写出一个最优性必要条件.

20. 已知(NLP)：

$$\min f(x)$$
$$\text{s.t} \ g_i(x) \leqslant 0, i=1,\cdots,m.$$

$\bar{x} \in X, I(\bar{x}) = \{i \mid g_i(\bar{x})=0, i=1,\cdots,m\}$. 设 $f(x), g_i(x), i \in I(\bar{x})$ 可微，$g_i(x), i \in I(\bar{x})$ 是凹函数，$g_i(\bar{x}), i \notin I(\bar{x})$ 连续. 研究下面的(LP)：

$$\min \langle \nabla f(\bar{x}), y \rangle$$
$$\text{s.t} \quad \langle \nabla g_i(\bar{x}), y \rangle \leqslant 0, i \in I(\bar{x}).$$
$$-1 \leqslant \eta_i \leqslant 1, i=1,\cdots,n, y=(\eta_1,\cdots,\eta_n).$$

如果 y^* 是(LP)的最优解，对应目标函数值为 Z^*.

1) 证明 $Z^* \leqslant 0$.

2）如果 $Z^*<0$，证明存在 $\delta>0$，使 $\bar{x}+\lambda y^*$ 对每个 $\lambda\in(0,\delta)$ 都可行，且 $f(\bar{x}+\lambda y^*)<f(\bar{x})$。

3）如果 $Z^*=0$，证明 $\bar{x}$ 满足 Kuhn-Tucker 必要条件.

21．设 $c\in R^n, b\in R^m$，A 是 $m\times n$ 矩阵，H 为 n 阶正定矩阵. 研究下面两个问题：

$$\text{(P 1)} \qquad \min\ \langle c,x\rangle+\frac{1}{2}\langle x,Hx\rangle$$
$$\text{s.t}\quad Ax\leqslant b.$$

$$\text{(P 2)} \qquad \min\ \langle h,y\rangle+\frac{1}{2}\langle y,Gy\rangle$$
$$\text{s.t}\quad y\geqslant\theta.$$

其中 $G=AH^{-1}A^*, h=AH^{-1}c+b$. 试讨论(P 1)和(P 2)的 Kuhn-Tucker 必要条件的关系.

22．已知：

$$\text{(P 1)} \qquad \min f(x)$$
$$\text{s.t}\quad g_i(x)\leqslant 0, i=1,\cdots,m,$$
$$x\in S.$$

$$\text{(P 2)} \qquad \min f(x)$$
$$\text{s.t}\quad \langle\lambda,(g_1(x),\cdots,g_m(x))\rangle\leqslant 0,$$
$$x\in S, \lambda=(\lambda_1,\cdots,\lambda_m).$$

试研究(P 1)和(P 2)的极小值点与 Kuhn-Tucker 必要条件的关系.

23．已知(NLP)：

$$\min f(x)=\langle c,x\rangle$$
$$\text{s.t}\quad g_i(x)\geqslant 0, i=1,\cdots,m,$$

其中 $c=(c_1,\cdots,c_n)\neq\theta$，$g_i(x)$ 可微. 如果 x^* 是极小值点，且在 x^* 一阶约束规格成立，证明 $I(x^*)\neq\varnothing$.

24．证明正切锥是闭锥.

25．在定理 2.13 中证明 $T(x)\subset G'(x)$.

26．研究(CP)：

$$\min f(x)$$
$$\text{s.t}\quad x\in S,$$

其中$f(x)$是R^n上的实值凸函数. $S\subset R^n$是非空凸集. $D(\bar{x})$是S在$\bar{x}$的可行方向锥. 证明$\bar{x}$是(CP)的极小值的充分必要条件是$f'(\bar{x}:d)\geqslant 0,\forall d\in D(\bar{x})$。

27. 证明推论3.3.1。

28. 证明定理3.4的3)、4)、5)。

29. 证明推论3.8.1。

30. 证明定理3.13.

31. 研究二次规划问题

$$\min \ \langle p,x\rangle+\frac{1}{2}\langle x,Cx\rangle$$

(QP)

$$\text{s.t} \qquad Ax\geqslant b.$$

其中C是n阶半正定矩阵，$p\in R^n$，A 是$m\times n$矩阵，$b\in R^m$. 证明x^*是(QP)的极小值点的充分必要条件是存在$u^*\in R^m$,满足：

$$C^*x+p-A^*u=\theta,$$

$$Ax\geqslant b,$$

$$\langle u,Ax\rangle=\langle u,b\rangle,$$

$$u\geqslant 0.$$

32. 设$f(x,y)$是R^{n+m}上的可微凸函数，$x\in R^n$, $y\in R^m$. A是$m\times n$矩阵，C是$l\times m$矩阵，$b\in R^m$，$d\in R^l$. 证明x^*,y^*是约束极值问题

$$\min f(x,y)$$

$$\text{s.t} \ \ Ax\geqslant b,$$

$$cy\geqslant d.$$

的极小值点的充分必要条件是：

1) x^*是约束极值问题

$$\min f(x,y^*)$$

$$\text{s.t} \ \ Ax\geqslant b.$$

的极小值点；

2) y^*是约束极值问题

$$\min f(x^*,y)$$

$$\text{s.t} \ \ Cx\geqslant d.$$

的极小值点。

33．研究(NLP)：

$$\min f(x)=\xi_1$$
$$\text{s.t } \xi_1^2+\xi_2^2=1, x=(\xi_1,\xi_2).$$

1) 写出对应的(DNLP)．证明目标函数是凹函数。

2) 求(NLP)和(DNLP)的解．比较它们的目标值。

34．已知下列(NLP)，分别求其(SP)的解：

1) min $\quad f(x)=(x-1)^2-2x$

s.t $\quad -x+1\leqslant 0, x\in R.$

2) min $\quad f(x)=-x$

s.t $\quad x^2\leqslant 0,$

$\quad x\geqslant 0, x\in R.$

3) min $\quad f(x)=x$

s.t $\quad x^3\leqslant 0,$

$\quad x\geqslant 0, x\in R.$

4) min $\quad f(x)=\xi_1+\xi_2$

s.t $\quad \xi_2-\xi_1^3\leqslant 0,$

$\quad \xi_2\geqslant 0, x=(\xi_1,\xi_2).$

35．设 $\psi(x,y)$ 是 x,y 的可微函数，$x\in R^n, y\in R^m$，$y\geqslant\theta, \bar{x},\bar{y}$ 满足：

$$\nabla\psi_x(x,y)=\theta,$$
$$\langle\nabla\psi_y(x,y),y\rangle=0,$$
$$\nabla\psi_y(x,y)\leqslant\theta, y\geqslant\theta.$$

$\psi(x,\bar{y})$ 是 x 的凸函数，$\psi(\bar{x},y)$ 是 y 的凹函数．证明 $\bar{x}$，$\bar{y}$ 是 $\psi(x,y)$ 的鞍点，$y\geqslant\theta$。

第五章 广义凸函数与广义凸规划

作为最后一章，我们讨论广义凸函数与广义凸规划问题.

近二十多年来，凸性理论已广泛应用到最优化的各个领域中. 而作为最优化的一个重要分支的数学规划,影响则更为深远. 但是在很多情况下，凸性对于数学规划的结果只是充分条件而远非必要条件. 同时，凸分析中很多重要性质的成立也不一定要求所讨论的函数是凸(凹)函数，而仅仅需要它的水平集是凸集就可以了. 所以我们将把水平集是凸集的函数作为这一章研究的起点.

广义凸性是数学规划研究的一个发展趋势. 因为研究凸集和凸函数的工具在推广数学规划的很多结果时常常已经足够了. 所以我们将直接推广以前学过的一些结果，其中许多类似的结论只列出来而将其证明略去. 有兴趣的读者、可以查阅书末所列的参考文献.

第一个研究其水平集是凸集的函数的学者是B.De Finetti[20],他没有称这类函数为拟凸函数,但是他发现这一类函数包括所有凸函数和一些非凸函数. W.Fenchel第一个称这类函数为拟凸函数，并且比较系统地研究了它的性质[17]. 较早地利用拟凸函数固有的结构推广凸函数的结果的是H.Nikaido 用 Brouwer的不动点定理推广Von Neumann极小极大原理[26]. 稍后，C.Berge 利用 Kakumani 定理修改了Nash和Sion的定理[10].

在最优化方面，M.Slater较早地推广了Kuhn-Tucker鞍点等价定理[20]. K.J.Arrow和A.C.Enthoven在 1961 年第一次发表处理拟凹最优化问题并应用到经济方面的文章[3]. 从六十年代末期开始,从事广义凸性研究的学者逐渐多起来,得到了一系列重

要的结果,其中比较突出的是B.Martos,L.Mangasarian, R.W.C-ottle, J.A.Ferland,J.P.Crouzeix和M.Auviel等人的工作.

§1 广义凸函数的定义和性质

1. 拟凸函数

定义1.1 设$f(x)$是R^n上的实值函数,如果对于每一对满足$f(x_2)\leqslant f(x_1)$的点x_1, $x_2\in R^n$, 有

$$f[(1-\lambda)x_1+\lambda x_2]\leqslant f(x_1),0\leqslant\lambda\leqslant1, \tag{1}$$

则称$f(x)$是拟凸函数. 如果$-f(x)$是拟凸函数,称$f(x)$是拟凹函数. 如果$f(x)$是拟凸函数也是拟凹函数,则称$f(x)$是拟仿射函数.

图1是R上的拟凸函数和拟凹函数的例子.

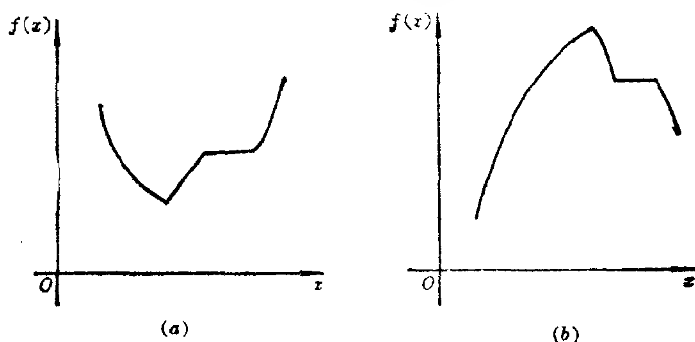


图1 R上的拟凸函数(a)和拟凹函数(b)

定理1.1 设$f(x)$是R^n上的实值函数,则下列结论彼此等价:

1) $f(x)$是拟凸函数.

2) 对于$\forall\alpha\in R$, 水平集 $S(f,\alpha)=\{x\in R^n|f(x)\leqslant\alpha\}$是凸集.

3) 在$f(x)$可微时,如果$f(x_2)\leqslant f(x_1)$,则

$$\langle\Delta f(x_1),\ x_2-x_1\rangle\leqslant0. \tag{2}$$

证　设 1) 成立. 对于 $\forall \alpha \in R$, $x_1, x_2 \in S(f,\alpha)$, 则

$$f(x_1) \leqslant \alpha,\ f(x_2) \leqslant \alpha.$$

因为$f(x)$是拟凸函数,故由定义 1.1,有

$$f[(1-\lambda)x_1+\lambda x_2] \leqslant \alpha,\ 0 \leqslant \lambda \leqslant 1. \tag{3}$$

这表示$(1-\lambda)x_1+\lambda x_2 \in S(f,\alpha)$,所以$S(f,\alpha)$是凸集,2) 成立.

设 2) 成立,即对于$\forall \alpha \in R$, $S(f,\alpha)$是凸集. 对于 $x_1, x_2 \in R^n$, 设$\bar{\alpha}=\max[f(x_1),f(x_2)]$,则$x_1, x_2 \in S(f,\bar{\alpha})$, 于是有

$$(1-\lambda)x_1+\lambda x_2 \in S(f,\bar{\alpha}),\ 0 \leqslant \lambda \leqslant 1.$$

所以有

$$f[(1-\lambda)x_1+\lambda x_2] \leqslant \bar{\alpha}=\max[f(x_1),f(x_2)],\ 0 \leqslant \lambda \leqslant 1, \tag{4}$$

故$f(x)$是拟凸函数,1) 成立. 所以 1) ,2) 等价.

设 1) 成立,则$f(x_2) \leqslant f(x_1)$时,有

$$f[(1-\lambda)x_1+\lambda x_2] \leqslant f(x_1),\ 0 \leqslant \lambda \leqslant 1. \tag{5}$$

但因为

$$f'(x_1:x_2-x_1)=\lim_{\lambda' \downarrow 0}\frac{f[x_1+\lambda'(x_2-x_1)]-f(x_1)}{\lambda'}$$
$$=\langle \nabla f(x_1), x_2-x_1 \rangle,$$

所以,有

$$\langle \nabla f(x_1), x_2-x_1 \rangle \leqslant 0,$$

即3) 成立.

设 3) 成立. 假设 1) 不成立, 即在 $f(x_2) \leqslant f(x_1)$时,存在 $x_3=(1-\lambda)x_1+\lambda x_2$, $0<\lambda<1$,有

$$f(x_1)<f(x_3). \tag{6}$$

由(2)知,这时有

$$\langle \nabla f(x_3), x_1-x_3 \rangle \leqslant 0, \tag{7}$$

$$\langle \nabla f(x_3), x_2-x_3 \rangle \leqslant 0, \tag{8}$$

于是可得

$$\langle \nabla f(x_3), x_2-x_1 \rangle \leqslant 0, \tag{9}$$

$$\langle \nabla f(x_3), x_1-x_2 \rangle \leqslant 0, \tag{10}$$

由(9),(10)式,得到

$$\langle \nabla f(x_3), x_1 - x_2 \rangle = 0. \tag{11}$$

定义

$$U = \{x \mid f(x) \leqslant f(x_1), x = \mu x_1 + (1-\mu)x_3, 0 \leqslant \mu \leqslant 1\}.$$

可以求出U中距x_3最近的点x_0. 由中值定理,在x_0, x_3之间存在$\bar{x}$,满足

$$f(x_3) = f(x_0) + \langle \nabla f(\bar{x}), x_3 - x_0 \rangle = f(x_0) + \mu_0 \langle \nabla f(\bar{x}), x_3 - x_1 \rangle, \tag{12}$$

其中 $0 < \mu_0 \leqslant 1$. 于是, 有

$$f(x_3) = f(x_0) + \lambda \mu_0 \langle \nabla f(\bar{x}), x_2 - x_1 \rangle. \tag{13}$$

同样,$\bar{x}$也是x_1,x_2的凸组合,且$f(x_1) < f(\bar{x})$. 进行类似于(6)-(11)的推证, 有

$$\langle \nabla f(\bar{x}), x_2 - x_1 \rangle = 0. \tag{14}$$

故由(13)式,$f(x_3) = f(x_0) \leqslant f(x_1)$. 这与(6)式矛盾. 所以 1)成立,1) ,3) 等价. ▌

由第二章定理 1.8 及其后的说明知,凸函数是拟凸函数,而拟凸函数不一定是凸函数.

例1 设$u(y)$是R^n到R的实值函数, $v(y)$是R^n到R^m的实值向量函数. 定义

$$C = \{x \geqslant 0, x \in R^m \mid \sup_y [u(y) \mid \langle x, v(y) \rangle \leqslant 0] < +\infty\},$$

$$f(x) = \sup_y \{u(y) \mid \langle x, v(y) \rangle \leqslant 0\}, x \in C.$$

D.G.Luenberger已经证明[22],C是R^m中的凸集, $f(x)$是C上的拟凸函数.

凸函数和拟凸函数之间有许多相似的性质, 现将其有关集合关系、连续性、可微性、有界性、极值、不等式等方面的性质列在后面,其中$S(f,\alpha)$表示水平集$\{x \mid f(x) \leqslant \alpha\}$,$H$表示$f(x)$的Hesse矩阵, 而称

$$D=\begin{bmatrix}0 & \nabla f(x)^T \\ \nabla f(x) & H\end{bmatrix}$$

是$f(x)$的加边Hesse矩阵.

性质 1 $f(x)$是凸函数的等价条件是epi f是凸集.

性质 1′ $f(x)$是拟凸函数的等价条件是 $S(f,\alpha)$ 是凸集，$\forall\alpha\in R$.

性质 1 和性质 1′是由Fenchel证明的. [17].

性质 2 $f(x)$是仿射函数的等价条件是 epi$f$,epi$(-f)$是凸集.

性质 2′ $f(x)$是拟仿射函数的等价条件是$S(f,\alpha)$,$S(-f,\alpha)$是凸集，$\forall\alpha\in R$.

性质 3 $f(x)$是仿射函数的等价条件是集合

$$\{(x,\mu)|f(x)=\mu,\mu\in R\}$$

是R^{n+1}中的凸集.

性质 3′ 设$f(x)$是拟仿射函数，则集合

$$U=\{x|f(x)=\mu,\mu\in R\}$$

是凸集. 反之,如果U是凸集,$f(x)$连续,则函数 $f(x)$是拟仿射函数.

在性质3′中，$f(x)$连续的条件是必不可少的.

例 2 在R中，设

$$f(x)=\begin{cases}0 & 0\leqslant x\leqslant 1,\\ 2 & 1<x\leqslant 2,\\ 1 & 2<x\leqslant 3,\\ +\infty & \text{其他}.\end{cases}$$

当$\mu=0$时,$U=[0,1]$. $\mu=1$时,$U=(2,3]$. $\mu=2$时,$U=(1,2]$. $\mu\neq 0,1,2$时,$U=\varnothing$. 可以检验,$f(x)$不是拟凸函数,而它在$x=1$,2 处发生间断(图 2).

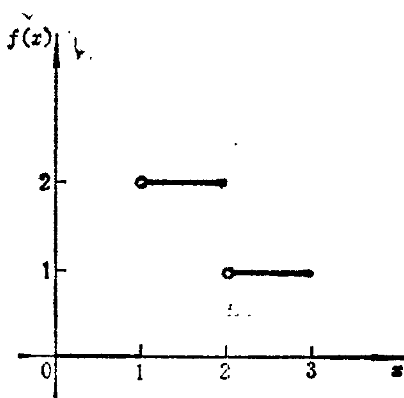


图 2

性质 4 $f(x)$是凸函数的等价条件是在$\lambda_i \geqslant 0, i=1,\cdots,m$, $\sum_{i=1}^{m} \lambda_i = 1$时，有

$$f(\lambda_1 x_1 + \cdots + \lambda_m x_m) \leqslant \lambda_1 f(x_1) + \cdots + \lambda_m f(x_m).$$

性质 4′ $f(x)$是拟凸函数的等价条件是在$\lambda_i \geqslant 0$, $i=1,\cdots,m$, $\sum_{i=1}^{m} \lambda_i = 1$ 时，有

$$f(\lambda_1 x_1 + \cdots + \lambda_m x_m) \leqslant \max\{f(x_1),\cdots,f(x_m)\}.$$

性质 5 如果$f(x)$是凸函数，则在$\mathrm{ri}(\mathrm{dom} f)$的每一个紧子集上$f(x)$有上界.

性质 5′ 如果$f(x)$是拟凸函数，则在 $\mathrm{ri}(\mathrm{dom} f)$的每一个紧子集上$f(x)$有上界.

性质 5′的证明见[36].

性质 6 如果 $f(x)$是凸函数，则对于 $\forall \alpha \in R$, $S(f,\alpha)$有界的等价条件是存在 α^*, $S(f,\alpha^*)$非空有界.

性质 6′ 设 $f(x)$是拟凸函数，如果 $S(f,\alpha)$非空有界，则存

在 $\alpha^*>\alpha$，$S(f,\alpha^*)$有界。

性质 6′的证明见[20]。

性质 7 如果 $f(x)$是凸函数，则 $f(x)$在 $\mathrm{ri}(\mathrm{dom} f)$上连续。

性质 7′ 如果 $f(x)$是拟凸函数，则 $f(x)$在 $\mathrm{ri}(\mathrm{dom} f)$ 上几乎处处连续。

性质 7′最先是由 Deak 证明的[20]。

性质 8 如果 $f(x)$是凸函数，则 $f'(x;y)$ 在 $\mathrm{ri}(\mathrm{dom} f)$ 上处处存在。

性质 8′ 如果 $f(x)$是拟凸函数，则 $f'(x;y)$ 在 $\mathrm{i}(\mathrm{dom} f)$上几乎处处存在。

性质 9 设 $f(x)$在 R^n 上二次连续可微，则 $f(x)$是凸函数的等价条件是 $f(x)$的 Hesse 矩阵 H 在 R^n 上半正定。

性质 9′ 设 $f(x)$在 R^n 上二次连续可微，$|D_j|$是 D 的 j 阶主子式，$j=0,1,\cdots,n$，$|D_0|=0$。如果 $f(x)$ 在 $R^n_+=\{x\in R_n|x\geqslant\theta\}$ 上是拟凸函数，则$|D_j|\leqslant 0$，$j=1,\cdots,n$。如果$|D_j|<0$，$j=1,\cdots,n$，则 $f(x)$在 R^n_+上是拟凸函数。

K. J. Arrow 和 A. C. Enthoven 以稍为不同的形式证明了性质 9′. [3]。

性质 10 设 $f(x)$在 R^n 上连续可微，则 $f(x)$是凸函数的等价条件是

$$f(y)-f(x)\geqslant\langle\nabla f(x),y-x\rangle.$$

性质 10′ 设 $f(x)$在 R^n 上连续可微，则 $f(x)$ 是拟凸函数的等价条件是当 $f(y)\leqslant f(x)$时，有

$$\langle\nabla f(x),y-x\rangle\leqslant 0.$$

性质 11 设 $f(x)$是 R^n 上的凸函数，则 $f(x)$的每一个局部极小值也是整体极小值。

性质 11′ 设 $f(x)$是 R^n 上的拟凸函数，则 $f(x)$的局部极小值是整体极小值，或在局部极小值点的一个领域内 $f(x)$是常

数.

在§2将证明性质11′.

性质 12 凸函数 $f(x)$的极小集是凸集.

性质 12′ 拟凸函数 $f(x)$的极小集是凸集.

性质12和性质12′是由凸函数和拟凸函数的水平集是凸集这个事实得来的,因为

$$\{x^*|f(x^*)\leqslant f(x),x\in R^n\}=\bigcap_{x\in R^n}\{x^*|f(x^*)\leqslant f(x)\}.$$

性质 13 设X、Y分别是R^n、R^m中的紧致子集,$f(x,y)$是从$X\oplus Y$到R的实值函数. 对每一个$y\in Y$,$f(x,y)$是x的凹函数. 对每一个$x\in X$,$f(x,y)$是y的凸函数. $f(x,y)$连续,则$f(x,y)$存在鞍点$(x^*,y^*)\in X\oplus Y$.

性质 13′ 设X、Y分别是R^n、R^m中的紧致子集,$f(x,y)$是从$X\oplus Y$到R的实值函数. 对每一个$y\in Y$,$f(x,y)$是x的拟凹函数且上半连续. 对每一个$x\in X$,$f(x,y)$是y的拟凸函数且下半连续,则$f(x,y)$存在鞍点$(x^*,y^*)\in X\oplus Y$.

性质13就是Von Neumann鞍点存在定理. 性质13′是它的推广, 其证明见[10].[20].

性质 14 设$f(x)$是R^n上的凸函数,$f(\theta)\leqslant 0$,则有

$$f(\lambda x)\leqslant\lambda f(x),0\leqslant\lambda\leqslant 1.$$

性质 14′ 设$f(x)$是R^n上的拟凸函数,$f(x)\geqslant f(\theta)$,则有

$$f(\lambda x)\leqslant f(x),0\leqslant\lambda\leqslant 1.$$

这个性质称为压缩性质.

性质 15 设$f(x)$是R^n上的凸函数,$f(\theta)=0$, $\lambda\in R$,则$g(\lambda)=f(\lambda x)/\lambda$在$\lambda>0$时是单调增加的.

性质 15′ 设$f(x)$是R^n上的拟凸函数,$f(x)\geqslant f(\theta)$,$\lambda\in R$,则$g(\lambda)=f(\lambda x)$在$\lambda\geqslant 0$时单调增加.

性质15是由Beckenbach和Bellman证明的. [20]. 性质15′作为其推广而直接得到.

性质 16 $f(x)$是 R^n 上的凸函数的等价条件是对于$\forall x, y$，$g(\lambda)=f[(1-\lambda)x+\lambda y]$是在$[0,1]$上的凸函数.

性质 16′ $f(x)$是 R^n 上的拟凸函数的等价条件是对于$\forall x, y$，$g(\lambda)=f[(1-\lambda)x+\lambda y]$是在$[0,1]$上的拟凸函数.

性质 16′的证明见[3].

性质 17 如果 $f_i(x)$是 R^n 上的凸函数，$i\in I$，I 是任意指标集，则 $\sup\limits_{i\in I} f_i(x)$是 R^n 上的凸函数.

性质 17′ 如果 $f_i(x)$是 R^n 上的拟凸函数，$i\in I$，I 是任意指标集，则 $\sup\limits_{i\in I} f_i(x)$ 是 R^n 上的拟凸函数.

性质 18 设 $f(x)$是 R^n 到$(-\infty,+\infty)$的凸函数，$F(y)$是 R 到$(-\infty,+\infty)$的不减少凸函数，则 $g(x)=F[f(x)]$ 是 R^n 上的凸函数.

性质 18′ 设 $f(x)$是 R^n 到$(-\infty,+\infty)$的拟凸函数，$F(y)$是 R 到$(-\infty,+\infty)$的不减少函数，则 $g(x)=F[f(x)]$是 R^n 上的拟凸函数.

性质 18′的重要性在于不需要函数 $F(y)$ 的凸性或拟凸性质.

在共轭对偶，次微分等方面还有许多相类似的性质，有兴趣的读者可以参考[11]，[13]，[36].

既然拟凸函数是凸函数的推广，所以有一些凸函数的性质对拟凸函数是不成立的.

1) 可加性. 对于 R^n 上的凸函数 $f(x)$，$g(x)$，当 $\alpha,\beta\geqslant 0$ 时，$\alpha f(x)+\beta g(x)$是凸函数.

对于拟凸函数，这个性质不再成立.

例 3 在R中，设

$$f_1(x)=\begin{cases}(x-1)^2 & 0\leqslant x\leqslant 2,\\ 1 & \text{其他}.\end{cases}$$

$$f_2(x)=\begin{cases}(x+1)^2 & -2\leqslant x\leqslant 0,\\ 1 & \text{其他}.\end{cases}$$

可以验证，$f_1(x)$，$f_2(x)$是拟凸函数，但

$$f_1(x)+f_2(x)=\begin{cases}(x-1)^2+1 & 0\leqslant x\leqslant 2.\\ (x+1)^2+1 & -2\leqslant x<0,\\ 2 & \text{其他}.\end{cases}$$

不是拟凸函数。

2) 下有界性. 对于 R^n 上的凸函数 $f(x)$，如果 $\mathrm{dom}f$ 有界，则 $\inf\limits_{x\in \mathrm{ri}(\mathrm{dom}f)} f(x)>-\infty$。

对于拟凸函数，这个性质不再成立。

例 4 在R中，设

$$f(x)=\begin{cases}\dfrac{1}{x-1} & 0\leqslant x<1,\\ 0 & x=1,\\ +\infty & \text{其它}.\end{cases}$$

容易验证 $f(x)$是拟凸函数，$\mathrm{dom}f=[0,1]$. 但

$$\inf_{x\in \mathrm{ri}(\mathrm{dom}f)} f(x)=-\infty。$$

2. 严格拟凸函数

定义 1.2 设 $f(x)$ 是 R^n 上的实值函数，如果当 $f(x_1)\neq f(x_2)$时有

$$f[(1-\lambda)x_1+\lambda x_2]<\max[f(x_1),f(x_2)],0<\lambda<1, \qquad (15)$$

则称 $f(x)$是严格拟凸函数.

严格拟凸函数不一定是拟凸函数，下面就是一个例子.

例 5 在R中，设

$$f(x)=\begin{cases}1 & x=0,\\ 0 & x\neq 0,\end{cases}$$

容易验证，$f(x)$是严格拟凸函数，但不是拟凸函数，因为$S(f,0)$

不是凸集.

Karamardian 已经证明，下半连续的严格拟凸函数是拟凸函数. [20].

定义 1.3 设 $f(x)$ 是 R^n 上的拟凸函数，如果 $f(x)$ 是严格拟凸函数，则称 $f(x)$ 是显拟凸函数.

显拟凸函数在半无限非线性规划中占有重要的地位，下面讨论它的一些性质.

1) 凸函数是显拟凸函数，直接利用定义 1.2 可以证明这一点.

2) 压缩性质：当 $f(x)$ 是显拟凸函数时，如果 $f(x)>f(\theta)$，则 $f(\lambda x)<f(x), 0<\lambda<1$.

3) 设 $f_i(x)$ 是 R^n 上的显拟凸函数，$i\in I$，I 是指标集. 逐点上确界函数 $f(x)=\sup\limits_{i\in I} f_i(x)$ 不一定是显拟凸函数.

例 6 在 R 中，设

$$f_i(x)=\begin{cases} x^3+i(x-10) & -1\leqslant x\leqslant 0, \\ i(x-10) & 0<x\leqslant 1, \\ (x-1)^2+i(x-10) & 1<x\leqslant 2, \\ +\infty & \text{其它}. \end{cases}$$

$I=\{i \mid i>0\}$. 可以验证，$f_i(x)$ 是显拟凸函数. 但是

$$f(x)=\sup_{i\in I} f_i(x)=\begin{cases} x^3 & -1\leqslant x\leqslant 0, \\ 0 & 0<x\leqslant 1, \\ (x-1)^2 & 1<x\leqslant 2, \\ +\infty & \text{其它} \end{cases}$$

不是显拟凸函数. 例如在 $x_1=-1, x_2=1, \lambda=\frac{1}{2}$ 时，$f(-1)=-1<f(1)=0$，而 $f(0)=0=\max[f(-1), f(1)]=0$.

在一定的条件下，显拟凸函数族的逐点上确界仍是显拟凸函数，下面的定理说明了这一点.

定理 1.2 设 $f_i(x)$是 R^n上的显拟凸函数，$i\in I$，I 是指标集．$\mathrm{dom} f_i=C$，$i\in I$，C 是 R^n 中的凸集．$f(x)=\sup\limits_{i\in I} f_i(x)$．如果对于每一个$x\in C$，至少存在一个 $i^*(x)\in I$，使$f(x)=f_{i^*}(x)$，则 $f(x)$是显拟凸函数．

证明 由上面的性质 17′，$f(x)$是拟凸函数，故只需证明$f(x)$是严格拟凸函数，即证明 $f(x_1)\neq f(x_2)$时，对于 $0<\lambda<1$，有

$$f[(1-\lambda)x_1+\lambda x_2]<\max[f(x_1),f(x_2)].$$

假设相反，在 $f(x_1)>f(x_2)$时，存在 $\lambda\in(0,1)$，有

$$f[(1-\lambda)x_1+\lambda x_2]=f(x_1). \tag{16}$$

设 $x=(1-\lambda)x_1+\lambda x_2$．由已知条件，在 I中存在 $i^*(x)=i^*$，$i^*(x_1)=i_1^*$，$i^*(x_2)=i_2^*$，满足

$$f(x)=f_{i^*}(x),f(x_1)=f_{i_1^*}(x_1),f(x_2)=f_{i_2^*}(x_2).$$

于是 $f_{i^*}(x)=f_{i_1^*}(x_1)$．因为 $f_{i^*}(x)$是拟凸函数，故

$$f_{i^*}(x)\leqslant\max\{f_{i^*}(x_1),f_{i^*}(x_2)\}. \tag{17}$$

当 $f_{i^*}(x_1)\neq f_{i^*}(x_2)$时，(17)式中的不等号是严格不等号．但由已知条件，知

$$f_{i^*}(x_1)\leqslant f_{i_1^*}(x_1),f_{i^*}(x_2)\leqslant f_{i_2^*}(x_2). \tag{18}$$

所以有

$$\begin{aligned}f(x)=f_{i^*}(x)&<\max\{f_{i^*}(x_1),f_{i^*}(x_2)\}\\&\leqslant\max\{f_{i_1^*}(x_1),f_{i_2^*}(x_2)\}\\&=\max\{f(x_1),f(x_2)\}.\end{aligned} \tag{19}$$

故 $f(x)<f(x_1)$，这和(16)式矛盾．

当 $f_{i^*}(x_1)=f_{i^*}(x_2)$时，由(17)，(18)式，有

$$f(x)\leqslant f_{i_2^*}(x_2)=f(x_2)<f(x_1),$$

与(16)式矛盾，所以 $f(x)$是严格拟凸函数．∎

在R上定义的拟凸函数，显拟凸函数具有一些特殊的性质，下面两个定理就是这方面的结果．

定理 1.3 设 $f(x)$ 是 R 上的实值函数，$\mathrm{dom} f = J$，则 $f(x)$ 是拟凸函数的充分必要条件是存在两个区间 $J^>$，$J^<$，$J^>$ 在 $J^<$ 的左边，$J = J^> \cup J^<$，$f(x)$ 在 $J^>$ 上不增加，在 $J^<$ 上不减少.

证明 充分性是明显成立的，仅证必要性. 设 $f(x)$ 是 R 上的拟凸函数. 令

$$J^> = \{x \in J \mid \text{存在 } y > x, \text{且 } f(x) > f(y)\}, \tag{20}$$

$$J^< = J \backslash J^>. \tag{21}$$

先证明 $f(x)$ 在 $J^>$ 上不增加，在 $J^<$ 上不减少.

假设相反，存在 $x_0 \in J^>$，$z < x_0$，而 $f(z) < f(x_0)$. 由 $J^>$ 的定义(20)式，存在 $y > x_0$，且 $f(x_0) > f(y)$. 这里，显然 x_0 是 y, z 的凸组合. 而

$$f(x_0) > \max\{f(y), f(z)\}.$$

故与 $f(x)$ 是拟凸函数矛盾，所以 $f(x)$ 在 $J^>$ 上不增加.

假设存在 $x_0 \in J^<$，$z > x_0$，$f(z) < f(x_0)$，由(20)式又有 $x_0 \in J^>$. 这与(21)式矛盾. 所以 $f(x)$ 在 $J^<$ 上不减少.

再证明 $J^>$，$J^<$ 是区间. 设 $x_0 \in J^>$，则由上面所证明，对于 $\forall z < x_0$，均有 $f(z) \geqslant f(x_0)$. 而由 $J^>$ 的定义知，存在 $y \in J$，$y > x$，有

$$f(y) < f(x_0) \leqslant f(z),$$

于是 $z \in J^>$，这表示 J 中位于 x_0 左边的所有 z 均属于 $J^>$，所以 $J^>$ 是一个区间. 由(21)式，$J^<$ 也是一个区间. ▌

定理 1.4 $f(x)$ 的条件与定理 1.3 相同，则 $f(x)$ 是显拟凸函数的充分必要条件是存在三个区间 $J^>$，$J^=$，$J^<$，$J = J^> \cup J^= \cup J^<$，$J^>$ 在 $J^=$ 的左边，$J^<$ 在 $J^=$ 的右边，$f(x)$ 在 $J^>$ 上严格单调减少，在 $J^=$ 上是常数，在 $J^<$ 上严格单调增加.

证明 充分性是明显成立的，仅证必要性. 设 $f(x)$ 是显拟凸函数. 令

$$\beta = \inf\{f(x) \mid x \in J\},$$

$$J^= = \{x \in J \mid f(x) \leqslant \beta\}, \tag{22}$$

$$J^> = J^> \setminus J^=, J^< = J^< \setminus J^=, \tag{23}$$

其中 $J^>, J^<$ 的意义与定理 1.3 中相同.

因为 $f(x)$ 是拟凸函数，故 $J^=$ 是凸集，从而是一个区间，$f(x)$ 在 $J^=$ 上显然是常数.

设 $x, y \in J^>, x < y$. 由定理 1.3, $J^>$ 是区间，故在 $\lambda \in [0.1]$ 上，区间 $\{(1-\lambda)x + \lambda y \mid 0 \leqslant \lambda \leqslant 1\} \subset J^>$. 但 $y \notin J^=$，所以 $f(y) > \beta$. 根据(23)式，$y \in J^>$，所以由(20)式，对 $\forall z < y$，有

$$f(z) \geqslant f(y) > \beta. \tag{24}$$

故 $\{(1-\lambda)x + \lambda y \mid 0 \leqslant \lambda \leqslant 1\} \cap J^= = \varnothing$, $J^>$ 在 $J^=$ 的左边. 由(23)式，$J^>$ 是一个区间.

因为 $J^> \subset J^>$，所以由定理 1.3，$f(x)$ 在 $J^>$ 上不增加. 但 $f(x)$ 是严格拟凸函数，不可能在 $J^>$ 的任何子区间上保持不变，所以 $f(x)$ 在 $J^>$ 上严格单调减少.

对于 $J^<$，可以进行类似的讨论，略. ▌

3. 强拟凸函数

定义 1.4 设 $f(x)$ 是 R^n 上的实值函数. 如果对于 $\forall x_1, x_2$, $x_1 \neq x_2$，有

$$f[(1-\lambda)x_1 + \lambda x_2] < \max\{f(x_1), f(x_2)\}, 0 < \lambda < 1, \tag{25}$$

则称 $f(x)$ 是强拟凸函数.

4. 伪凸函数与严格伪凸函数

定义 1.5 设 $f(x)$ 是 R^n 上的可微实值函数. 如果在 $f(x_2) < f(x_1)$ 时，有

$$\langle \nabla f(x_1), x_2 - x_1 \rangle < 0, \tag{26}$$

则称 $f(x)$ 是伪凸函数. 如果在 $f(x_2) \leqslant f(x_1)$ 时，有

$$\langle \nabla f(x_1), x_2 - x_1 \rangle < 0, \tag{27}$$

则称 $f(x)$ 是严格伪凸函数.

当 $-f(x)$ 是(严格)伪凸函数时，称 $f(x)$ 是(严格)伪凹函数.

图 3 是 R 中的伪凸函数与伪凹函数的图形.

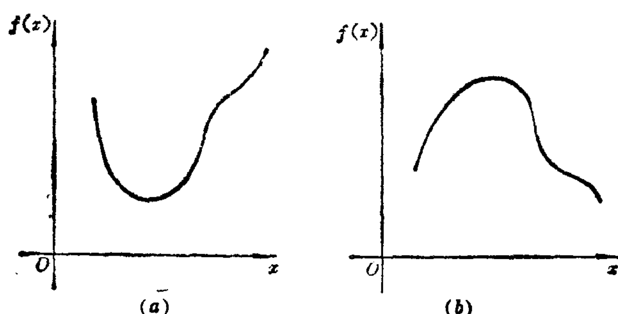


图 3 R 中的伪凸函数(a)和伪凹函数(b)

在定义 1.5 中也可以不要求 $f(x)$ 的可微性来定义伪凸函数，其具体内容见[37].

5. 七类凸函数、广义凸函数之间的关系

为了讨论方便，采用下列记号：

严格凸函数：SC；凸函数：C；严格伪凸函数：SPC；伪凸函数：PC；强拟凸函数：STQC；严格拟凸函数：SQC；拟凸函数：QC.

在下面的定理中，设 $f(x)$ 是连续可微的.

定理 1.5 七类凸函数、广义凸函数具有下面的关系：SC⇒C，C⇒PC，PC⇒SQC，SQC⇒QC，SC⇒STQC，STQC⇒SQC，PC + STQC⇒SPC.

证明 仅证 PC⇒SQC. 其余几个结论的证明是类似的.

假设 $f(x)$ 是 PC，而不是 SQC. 由定义 1.2，存在 x_1，x_2，$f(x_2)<f(x_1)$，而对某一个 $\hat{x}=(1-\lambda)x_1+\lambda x_2, 0<\lambda<1$，有

$$f(\hat{x})\geqslant f(x_1).$$

故存在一个 $\bar{\lambda}, 0<\bar{\lambda}<1, \bar{x}=(1-\bar{\lambda})x_1+\bar{\lambda}x_2$，且

$$f(\bar{x})=\max\{f(x)\,|\,x=(1-\lambda)x_1+\lambda x_2, 0\leqslant\lambda\leqslant 1\}. \tag{28}$$

所以容易证明

$$\langle \nabla f(\bar{x}), x_1-\bar{x}\rangle \leqslant 0, \tag{29}$$

$$\langle \nabla f(\bar{x}), x_2-\bar{x}\rangle \leqslant 0. \tag{30}$$

由(29)式，有

$$0 \geqslant \langle \nabla f(\bar{x}), x_1-\bar{x}\rangle = \bar{\lambda}\langle \nabla f(\bar{x}), x_1-x_2\rangle. \tag{31}$$

由(30)式，有

$$0 \geqslant \langle \nabla f(\bar{x}), x_2-\bar{x}\rangle = (1-\bar{\lambda})\langle \nabla f(\bar{x}), x_2-x_1\rangle \tag{32}$$

故利用(31)、(32)式及 $\bar{\lambda}>0, 1-\bar{\lambda}>0$ 得到

$$\langle \nabla f(\bar{x}), x_1-x_2\rangle = 0. \tag{33}$$

从而由(32)式，有

$$\langle \nabla f(\bar{x}), x_2-\bar{x}\rangle = 0. \tag{34}$$

因为 $f(x)$ 是 PC，由(34)式知 $f(x_2) \geqslant f(\bar{x})$，故 $f(x_1) > f(\bar{x})$. 这与(28)式矛盾. 所以 $f(x)$ 是 SQC. ▌

定理 1.5 中的各个关系的逆关系一般都不成立. 下面的表中用一些函数作为例子说明上述的关系. 其中“+”表示“是”，“—”表示“非”.（见下页）.

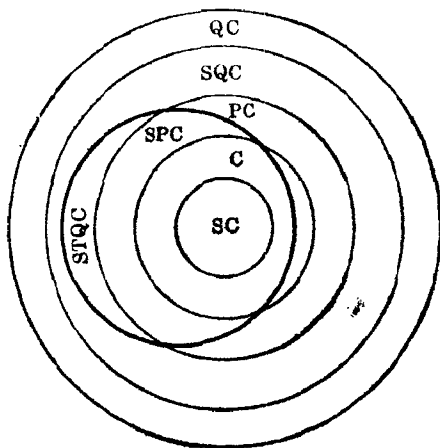


图 4

图 4 是上述各类凸函数、广义凸函数之间的关系的图示，从中也可以看出，拟凸函数凸性最弱，而严格凸函数凸性最强．强拟凸函数与伪凸函数，凸函数之间不存在包含关系．

表一　一些函数的凸性和广义凸性　$x,y\in R$

	SC	C	SPC	PC	SQC	QC	STQC
$f_1(x)=\begin{cases}0 & 0\leqslant x\leqslant 1,\\ -(x-1)^2 & 1<x\leqslant 2.\end{cases}$	−	−	−	−	−	+	−
$f_2(x,y)=-x^2,\ x\geqslant 0,y\leqslant 1.$	−	−	−	−	+	+	−
$f_3(x,y)=-x^2-4,\ x\geqslant 0,y\leqslant 1.$	−	−	−	+	+	+	−
$f_4(x)=0,\ \ 0\leqslant x\leqslant 1.$	−	−	−	+	+	+	−
$f_5(x)=-x,\ 0\leqslant x\leqslant 1.$	−	+	+	+	+	+	+
$f_6(x)=x^2,\ 0\leqslant x\leqslant 1.$	+	+	+	+	+	+	+
$f_7(x)=-x^2-x,\ 0\leqslant x\leqslant 1.$	−	−	+	+	+	+	+
$f_8(x)=-x^2,\ 0\leqslant x\leqslant 1.$	−	−	−	−	+	+	+

§2　广义凸规划

我们已经知道，对于凸函数，局部极小值一定是整体极小值，但是对于广义凸函数，这个性质就不一定成立．一个拟凸函数的局部极小值就可能不是整体极小值，见图 5．这一节将讨论局部极小值与整体极小值之间的关系及广义凸规划的最优性条件．

1．局部极小值与整体极小值的关系

定理 2.1　设 $f(x)$是 R^n上的拟凸函数．如果 x^*是 $f(x)$的严格局部极小值点，则 x^* 也是 $f(x)$的严格整体极小值点．

证明　设 x^* 是严格局部极小值点，即存在$\delta>0$，当 $x\in N_\delta(x^*)=\{x\,|\,|x-x^*|\leqslant\delta\}$时，

$$f(x)>f(x^*).\tag{1}$$

如果 x^* 不是严格整体极小值点，即存在 $\bar{x}$，$\bar{x}\neq x^*$，且 $f(\bar{x})<f(x^*)$．由 $f(x)$的拟凸性知

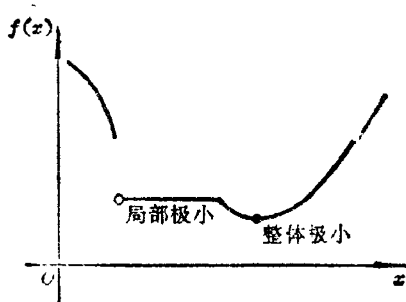


图 5

$$f[\lambda\bar{x}+(1-\lambda)x^*]\leqslant f(x^*),0<\lambda<1. \tag{2}$$

而 $\lambda\to 0$ 时，$\lambda\bar{x}+(1-\lambda)x^*\in N_\delta(x^*)$，这与(1)矛盾. ▌

定理 2.2 如果拟凸函数 $f(x)$ 的局部极小值点 x_1 不是整体极小值点，则存在 x_1 的一个邻域，$f(x)$ 在这个邻域和连结 x_1 与任意整体极小值点 x_2 的线段的交上是一个常数.

证明 假设不存在这样的邻域. 因为 x_1 是 $f(x)$ 的局部极小值点，故一定存在 $x_3=\lambda x_1+(1-\lambda)x_2$，使 $f(x_3)>f(x_1)$. 另一方面，由 $f(x)$ 的拟凸性，又有

$$f(x_3)\leqslant f(x_1),$$

矛盾. ▌

对于严格拟凸函数，情形就不同了.

定理 2.3 严格拟凸函数 $f(x)$ 的局部极小值点一定是整体极小值点.

证明 假设 x_1 是 $f(x)$ 的局部极小值点而不是整体极小值点，则存在 x_2，满足

$$f(x_2)<f(x_1).$$

由 $f(x)$ 的严格拟凸性，这表示

$$f[\lambda x_1+(1-\lambda)x_2]<f(x_1),0<\lambda<1. \tag{3}$$

令 $\lambda\to 1$，对充分接近 x_1 的 x_3，由(3)式有

$$f(x_2)<f(x_1),$$

这与 x_1 是局部极小值点矛盾. ▌

定理 2.4 强拟凸函数不在两个点上达到极小值.

证明留给读者.

定理 2.5 设 $f(x)$是 R^n 上的伪凸函数, dom$f$ 是开集, x^* 是其一个稳定点,即 $\nabla f(x^*)=\theta$,则 x^*是 $f(x)$的整体极小值点.

证明 因为对任意 x, 均有

$$\langle x-x^*,\nabla f(x^*)\rangle=0. \tag{4}$$

于是由 $f(x)$的伪凸性,知

$$f(x)\geqslant f(x^*),$$

即 x^*是整体极小值点. ▌

定理 2.6 设 $f(x)$是 R^n上的严格伪凸函数,dom$f$ 是开集, x^*是其稳定点,则 x^*是 $f(x)$的唯一整体极小值点.

证明留给读者.

下面将以上六个定理的结果列成一个表, 读者不难看出极小值点的性质与凸性有关.

	严格局部极小值点是严格整体极小值点	局部极小值点是整体极小值点	局部极小值点是唯一整体极小值点	稳定点是整体极小值点	稳定点是唯一整体极小值点
拟凸函数	+				
严格拟凸函数	+	+			
强拟凸函数	+	+	+		
伪凸函数	+	+	+	+	
严格伪凸函数	+	+	+	+	+

2. 广义凸规划中的最优性条件

如果 $f(x),g_i(x),i=1,\cdots,m,h_j(x),j=1,\cdots,p$,是广义凸函数,则称约束极值问题

$$\min f(x)$$

$$\text{s.t. } g_i(x)\leqslant 0, i=1,\cdots,m, \tag{5}$$

(GCP) $$h_j(x)=0, j=1,\cdots,p, \tag{6}$$

$$x\in S.$$

是广义凸规划(GCP)．其中 S 是凸集．

下面讨论(GCP)的最优性条件，X 表示可行解集．

定理 2.7 设 $f(x), g_i(x), i=1,\cdots,m, h_j(x), j=1,\cdots,p$，是 R^n 上的实值函数，在开凸集 S 上可微，$f(x)$ 是伪凸函数，$g_i(x^*)$，$i\in I(x^*)$ 是拟凸函数．$h_j(x), j=1,\cdots,p$，是拟仿射函数，如果存在 $\lambda_1,\cdots,\lambda_m,\mu_1,\cdots,\mu_p$ 满足

$$\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)+\sum_{j=1}^{p}\mu_j\nabla h_j(x^*)=\theta, \tag{7}$$

$$\lambda_i g_i(x^*)=0, \tag{8}$$

$$\lambda_i\geqslant 0, i=1,\cdots,m,$$

则 $x^*\in X$ 是(GCP)的解，其中 $I(x^*)=\{i\,|\,g_i(x^*)=0, i=1,\cdots,m\}$．

证明 任取 $x\in X$，则由

$$g_i(x)\leqslant g_i(x^*)=0, i\in I(x^*),$$

$$h_j(x)=h_j(x^*), j=1,\cdots,p,$$

知

$$\langle x-x^*,\nabla g_i(x^*)\rangle\leqslant 0, i\in I(x^*), \tag{9}$$

$$\langle x-x^*,\nabla h_j(x^*)\rangle=0, j=1,\cdots,p. \tag{10}$$

如果在 $i\notin I(x^*)$ 时令 $\lambda_i=0$，则由(9)，(10)式有

$$\langle x-x^*, \sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)+\sum_{j=1}^{p}\mu_j\nabla h_j(x^*)\rangle\leqslant 0.$$

故由(7)式知

$$\langle x-x^*,\nabla f(x^*)\rangle\geqslant 0. \tag{11}$$

但由 $f(x)$ 的伪凸性，知 $f(x)\geqslant f(x^*)$. ∎

定理 2.8 设 $f(x), g_i(x), i=1,\cdots,m$，是 R^n 上的实值可微拟凸函数，$h_j(x), j=1,\cdots,p$，是 R^n 上的实值可微拟仿射函数，如果存在 $x^*\in R^n, \lambda_i\geqslant 0, i=1,\cdots,m, \mu_j, j=1,\cdots,p$，满足(7)，(8)式，且

1) 存在 $x_1\in R^n$，满足

$$\langle x_1-x^*, \nabla f(x^*)\rangle>0;$$

或

2) $\nabla f(x^*)\neq\theta$，$f(x)$ 二阶可微，

则 x^* 是(GCP)的解.

定理 2.8 的证明见[18].

在更广泛的一类广义凸规划中，例如目标函数 $f(x)$ 是 PCN 的，约束函数 $g_i(x)$ 是 QCN 的，[5]、[36] 讨论了更一般的最优性条件及广义对偶凸规划.

§3 拟凸函数和伪凸函数的判别准则

随着对广义凸函数的深入研究，已经从各个角度对它们的判别准则进行了讨论. 本节介绍一些主要结果.

1. 一阶条件

在§2 的定理 2.1-定理 2.6 中，已经证明在各种广义凸性的假设下，局部极小与整体极小、稳定点与整体极小的关系. 反过来，由这些关系却不一定能推出它们应属于哪一类广义凸函数. 但是，在拟凸函数的前提下，由它们则可以得到一定的结果.

定理 3.1 设 $f(x)$ 是 R^n 上的拟凸函数，如果 x^* 是 $f(x)$ 的局部极小值点时，x^* 就是 $f(x)$ 的整体极小值点，则 $f(x)$ 是严格拟凸函数.

定理 3.2 设 $f(x)$ 是 R^n 上的拟凸函数，如果 $f(x)$ 只有唯一的整体极小值点，则 $f(x)$ 是强拟凸函数.

定理 3.3 设 $f(x)$是 R^n 上的可微拟凸函数，如果 $f(x)$ 的唯一稳定点也是它的整体极小值点，则 $f(x)$是严格伪凸函数.

以上三个定理的证明见[37].

定理 3.4 设 $f(x)$是 R^n 上的可微拟凸函数. $\mathrm{dom} f=S$ 是开集. 如果$\nabla f(x^*)=\theta$时，x^*就是 $f(x)$的整体极小值点，则 $f(x)$是伪凸函数.

证明 假设相反，$f(x)$不是伪凸函数，则存在 $x,y\in S$，满足

$$f(y)<f(x), \tag{1}$$

$$\langle y-x,\nabla f(x)\rangle\geqslant 0. \tag{2}$$

故$\nabla f(x)\neq\theta$. 否则 x 是 $f(x)$的整体极小值点，与(1)式矛盾.

因为 $f(x)$是拟凸函数，由(1)式知

$$\langle y-x,\nabla f(x)\rangle\leqslant 0. \tag{3}$$

结合(2)式，得$\langle y-x,\nabla f(x)\rangle=0$.

由 $f(x)$的可微性知它在 S 上连续，故存在 $\rho>0$，使 $f[y+\rho\nabla f(x)]<f(x)$.

由 $f(x)$的拟凸性，在 $t\in(0,1)$时

$$\frac{f\{x+t[y+\rho\nabla f(x)-x]\}-f(x)}{t}\leqslant 0. \tag{4}$$

在(4)中令 $t\to 0$，得

$$\begin{aligned}&\langle y+\rho\nabla f(x)-x,\nabla f(x)\rangle\\=&\rho\langle\nabla f(x),\nabla f(x)\rangle\leqslant 0.\end{aligned} \tag{5}$$

而$\nabla f(x)\neq\theta$，$\rho>0$，矛盾. ▌

定理 3.5 设 $f(x)$是 R^n 上的可微拟凸函数，$\mathrm{dom} f=S$是开集，如果$\nabla f(x^*)=\theta$时，x^*就是 $f(x)$ 的局部极小值点，则 $f(x)$是伪凸函数. 它的逆也成立.

证明 由定理 3.4，只要证明$\nabla f(x^*)=\theta$ 时，x^*是 $f(x)$的整体极小值点即可. 假设存在 $\bar{x}\in S$，$\nabla f(\bar{x})=\theta$，而 $\bar{x}$ 不是整体极小值点. 设

$$T=\{x\in S \mid f(x)<f(\bar{x})\},$$

则 T 是 S 的非空凸子集.

因为 $f(\bar{x})$ 是局部极小值，故 $\bar{x}\notin \mathrm{cl}T$. 设

$$K=\{x\in S \mid x=\bar{x}-t(y-\bar{x}),\ t\in(0,1),\ y\in T\},$$

于是 K 是 S 的开凸子集，而由 $f(x)$ 的拟凸性，对于 $\forall x\in K$，有 $f(x)\leqslant f(\bar{x})$.

对于 $\forall y\in T$，总存在 $t_y\in(0,1)$，满足

$$\bar{x}+t(y-\bar{x})\notin T,\ t\in[0,t_y],$$

$$\bar{x}+t(y-\bar{x})\in T,\ t\in(t_y,1].$$

设 $z=\bar{x}+t_y(y-\bar{x})$，则由 $f(x)$ 在 z 的连续性，知 $f(z)=f(\bar{x})$. 由前设，K 是 z 的一个邻域，且对于 $\forall x\in K$，$f(x)\leqslant f(z)$. 则由 $f(x)$ 在 z 的可微性知 $\nabla f(z)=\theta$. 但 z 不是 $f(x)$ 的局部极小值点，矛盾. ▎

上述证明示意图见图 6.

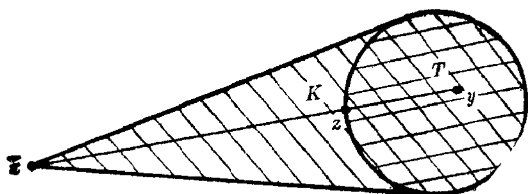


图 6

2. 二阶条件

一般来说，二阶条件是从两个方面给出的. 一类是用矩阵的形式，即函数的 Hesse 矩阵具有某种性质而判断广义凸性. 另一类是用加边行列式的某种性质而判断广义凸性.

定理3.6 设 $f(x)$ 是 R^n 上的二阶连续可微的实值函数，$\mathrm{dom}f=S$ 是开凸集，则 $f(x)$ 是拟凸函数的充分必要条件是：

1) $x\in S,h\in R^n,\langle\nabla f(x),h\rangle=0$ 时，有

$$\langle h, H(x)h\rangle \geqslant 0, \tag{6}$$

其中 $H(x)$ 是 $f(x)$ 在 x 的 Hesse 矩阵.

2) 当 $\nabla f(x)=\theta$ 时,对于 $\forall h\in R^n$, 函数 $f_{x,h}(t)=f(x+th)$ 在 $\{t \mid x+th\in S, t\in R\}$ 上是拟凸函数.

这个定理是 J.P.Crouzeix 在[13]中证明的.

注1 当定理 3.6 中的 2)用下面两个条件之一代替时,就和1)一起构成判断拟凸函数的充分条件[13]:

2*) 当 $\nabla f(x)=\theta$ 时,对每一个 $h\in R^n$, $h\neq\theta$, 有

$$\langle h, H(x)h\rangle > 0.$$

2**) 当 $\nabla f(x)=\theta$ 时, x 是 $f(x)$ 的整体极小值点.

注2 可以证明(见[14]、[13]),定理 3.6 中的 1) 和下列各条件之一等价:

1*) $H(x)$ 半正定,或 $H(x)$ 有一个负特征根且存在一个 $h\in R^n$, 满足

$$H^*(x)h=\nabla f(x), \langle \nabla f(x), h\rangle \leqslant 0.$$

1**) 加边 Hesse 矩阵

$$D=\begin{bmatrix} 0 & \nabla f(x)^T \\ \nabla f(x) & H(x) \end{bmatrix}$$

有一个负特征根,其中 $\nabla f(x)^T$ 是 $\nabla f(x)$ 的转置.

1***) 对于 $\{1,\cdots,n\}$ 的所有非空子集 P, 行列式

$$|D_P|=\begin{vmatrix} 0 & \nabla f(x)_P^T \\ \nabla f(x)_P & H(x)_P \end{vmatrix} \leqslant 0,$$

其中 $H(x)_P$ 是由 $H(x)$ 划去行标、列标不在 P 中的那些行和列得到的. $\nabla f(x)_P$、$\nabla f(x)_P^T$ 的取法相似.

定理 3.7 设 $f(x)$ 是 R^n 上的二阶连续可微实值函数, $\mathrm{dom} f=S$ 是开凸集. 行列式

$$|D|=\begin{vmatrix} 0 & \nabla f(x)^T \\ \nabla f(x) & H(x) \end{vmatrix}$$

称为 $f(x)$的加边行列式. 如果在 S 上$|D|$的顺序主子式

$$|D_k|=\begin{vmatrix} 0 & f'_{\xi_1} & \cdots & f'_{\xi_k} \\ f'_{\xi_1} & f''_{\xi_1^2} & \cdots & f''_{\xi_1\xi_k} \\ \cdots & \cdots & \cdots & \cdots \\ f'_{\xi_k} & f''_{\xi_k\xi_1} & \cdots & f''_{\xi_k^2} \end{vmatrix}<0, k=1,\cdots,n, \tag{7}$$

则 $f(x)$是拟凸函数.

这个定理的证明见[8].

定理 3.8 设 $f(x)$是 R^n上的二阶可微实值函数,且$\mathrm{dom}f=S$是开凸集,则 $f(x)$是伪凸函数的充分必要条件是:

1) $x\in S, h\in R^n, \langle\nabla f(x),h\rangle=0$ 时

$$\langle h^T H(x),h\rangle\geqslant 0.$$

2) $\nabla f(x)=\theta$ 时,x 是 $f(x)$的局部极小值点.

证明 必要性. 设 $f(x)$是伪凸函数,则 $f(x)$是拟凸函数. 由定理 3.6,1)成立. 而由定理 3.5,只要 $\nabla f(x)=\theta$, x 就是 $f(x)$的局部极小值点.

充分性.由定理3.5,只要由 1)、2)证明$f(x)$是拟凸函数即可.

假设相反,$f(x)$不是拟凸函数,则存在 $x,y\in S, x\neq y$,满足

$$M=\sup_{t\in[0,1]} f[x+t(y-x)]>\max\{f(x),f(y)\}. \tag{8}$$

设 $\varphi(t)=f[x+t(y-x)]$,则(8)式可以记为

$$M=\sup_{t\in[0,1]}\varphi(t)>\max\{\varphi(0),\varphi(1)\}. \tag{9}$$

因为 $\varphi(t)$在$[0,1]$连续,存在某一个数 $t_0\in(0,1)$,使

$$\varphi(t)<M,\ 0\leqslant t<t_0.$$

显然,$\varphi'(t_0)=\langle h,\nabla f(x+t_0h)\rangle=0$,其中 $h=y-x$. 由 t_0 的定义及条件 2)知$\nabla f(x+t_0h)\neq\theta$.

不失一般性,设 $x+t_0h=a=\theta, f(a)=0$, 而$\nabla f(a)=-e_n$,e_n是标准正交基的第 n 个向量.将$x=(\xi_1,\cdots,\xi_n)$表示成$x=(r,\lambda)$,其中 $r=(\xi_1,\cdots,\xi_{n-1}), \lambda=\xi_n$. 而由 $\langle\nabla f(a),h\rangle=0$ 知

$$h=(u,0), \qquad u\in R^{n-1}.$$

但$\nabla f(\theta,0)=(\theta,-1)$，于是由隐函数定理知存在 R^{n-1}中原点的邻域U及U中的实函数 $g(r)$，在 $r\in U$时满足

$$f[r,g(r)]=0, f'_\lambda[r,g(r)]<0, \tag{10}$$

$$\nabla g(r)=-\frac{1}{f'_\lambda[r,g(r)]}\nabla f_r[r,g(r)], \tag{11}$$

$$\begin{aligned} g''(r)=\frac{-1}{f'_\lambda[r,g(r)]}[&f''_{rr}(r,g(r)) \\ &+\nabla g(r)f'_{\lambda r}(r,g(r))^T+f''_{r\lambda}(r,g(r))\nabla g(r)^T \\ &+f''_{\lambda\lambda}(r,g(r))\langle\nabla g(r),\nabla g(r)\rangle], \end{aligned} \tag{12}$$

所以 $g(\theta)=0,\nabla g(\theta)=\theta$.

可以证明 $g(r)$是凸函数且 $g(tu)>0, tu\in U, t\in[0,t_0)$.

由$f'(x)$的连续性，知存在 R^{n-1}中原点的邻域$V\subset U$及$\xi>0$，当$(r,\lambda)\in V\oplus[-\xi,\xi]$时

$$f'_\lambda(r,\lambda)<0. \tag{13}$$

最后由 $g(r)$ 的连续性，存在$t\in[0,t_0)$，使$tu\in V$及$0<g(tu)<\xi$. 因而

$$f(tu,0)<0, \tag{14}$$

$$f[tu,g(tu)]=0. \tag{15}$$

于是$\nabla f(tu,\lambda)\geqslant\theta$.而由(13)，对于$\lambda\in[0,g(tu)]$，有

$$\langle e_n,\nabla f(tu,\lambda)\rangle<0,$$

故矛盾. 所以 $f(x)$是拟凸函数. ▌

注 1 W. E. Diewert, M. Avriel 和 I. Zang已经证明（见[14]）定理中的2)可以用下面的条件代替：

2*)对于满足 $\langle\nabla f(x), h\rangle=0$ 及 $\langle h^TH(x), h\rangle=0$ 的所有 h，函数 $\phi(t)=f(x+th)$在 $t=0$ 达到关于$\{t\mid x+th\in S\}$ 的局部极小值.

注 2 定理3.6注2中的三个条件分别与定理3.8中1)等

价.

注 3 定理中的 2)用下面的条件代替时：

2**) 只要$\nabla f(x)=\theta$,就存在 x 的一个凸邻域 $V\subset S$,满足

$$\langle (x-y)^T H(y), x-y\rangle \geqslant 0, y\in V.$$

这时 1)和 2**)是 $f(x)$是伪凸函数的充分条件.

下面两个定理利用加边行列式 $|D(x)|$ 和 Hesse 矩阵行列式 $|H(x)|$判断伪凸性,其中$|D_P(x)|$,$|H_P(x)|$与定理 3.6 注 3 的 1***)中的定义相同.

定理 3.9[6] 设 $f(x)$是 R^n 上的二阶连续可微实值函数, $\mathrm{dom} f=S$ 是开凸集. 如果：

1) 对于$\{1,\cdots,n\}$的所有非空子集P,$|D_P(x)|\leqslant 0$.

2) 若对于某个 $\bar{x}\in S$,$D_P(\bar{x})=0$,那么在 $\bar{x}$ 的一个邻域中的所有 x,$|H_P(x)|\geqslant 0$.

则 $f(x)$是伪凸函数.

定理 3.10[14] 设 $f(x)$是 R^n上的二阶连续可微实值函数, $\mathrm{dom} f=S$ 是开凸集. 如果：

1) 对于$\{1,\cdots,n\}$的所有非空子集P,$|D_P(x)|\leqslant 0$,

2) 只要 $\nabla f(\bar{x})=\theta$,就存在 $\bar{x}$ 的邻域$V\subset S$,使对于 $\forall x\in V$, $P\subset\{1,\cdots,n\}$,有 $|H_P(x)|\geqslant 0$.

则 $f(x)$ 是伪凸函数

M. Avriel 和 S. Schaible 等人利用广义 Hesse 矩阵讨论伪凸性,得到了不少有意义的结果[6],下面是其中一个结论.

定义 3.1 设 $f(x)$是 R^n上的二阶可微实值函数,$H(x)$是其 Hesse 矩阵. 称

$$H(x;r(x))=H(x)+r(x)\nabla f(x)\nabla f(x)^T$$

是 $f(x)$的广义 Hesse 矩阵,其中 $r(x)$是 R^n上的实值函数。

定理 3.11 如果存在连续函数 $r(x)$,使$H(x;r(x))$半正定,则 $f(x)$是伪凸函数.

3．二次函数的情形

由于二次规划在数学规划中的重要性，这一小节专门叙述关于二次函数与前面平行的一些结果.

二次函数是指

$$q(x)=\frac{1}{2}x^TQx+a^Tx, \tag{16}$$

其中 $x,a\in R^n$，Q 是 n 阶实对称矩阵. 对于(16)，它的 Hesse 矩阵就是 Q，而加边 Hesse 矩阵是

$$D(x)=\begin{bmatrix} 0 & (Qx+a)^T \\ Qx+a & Q \end{bmatrix} \tag{17}$$

下面的结果可以在[6]、[12]、[18]、[30]中找到.

定理 3.12 二次函数 $q(x)$ 是凸函数的等价条件是它是拟凸函数.

定义 3.2 如果 $f(x)$ 是拟凸(伪凸)函数而不是凸函数，则称 $f(x)$ 是仅拟凸(仅伪凸)函数.

定理 3.13 二次函数 $q(x)$ 是仅拟凸函数的等价条件是 Q 和 $D(x)$ 都只有一个负特征根. $q(x)$ 在非负真锥 R^n_+ 上是仅拟凸函数的等价条件是二次函数

$$\Phi(x,\xi)=\frac{1}{2}\begin{pmatrix} x \\ \xi \end{pmatrix}^T\begin{pmatrix} Q & a \\ a^T & 0 \end{pmatrix}\begin{pmatrix} x \\ \xi \end{pmatrix} \tag{18}$$

在非负真锥 R^{n+1}_+ 上是拟凸函数，其中 $\xi\in R$，$a\in R^n$.

定理 3.14 如果二次函数 $q(x)$ 在非负真锥上是拟凸函数且 $a\neq\theta$，则 $q(x)$ 在非负真锥上也是伪凸函数.

定理 3.15 二次函数 $q(x)$ 在凸集 C 上是伪凸函数的等价条件是对于 $\{1,\cdots,n\}$ 的所有非空子集 P，$|D_P(x)|\leqslant 0$，$x\in C$. 如果 $|D_P(x)|=0$，则 $|H_P(x)|=|Q_P|\geqslant 0$.

定理 3.16 二次函数 $q(x)$ 在不包含原点的非负真锥上是伪凸函数的等价条件是二次函数

$$\Phi(x,\xi)=\frac{1}{2}\begin{pmatrix} x \\ \xi \end{pmatrix}^T \begin{pmatrix} Q & a \\ a^T & 0 \end{pmatrix} \begin{pmatrix} x \\ \xi \end{pmatrix} \tag{19}$$

在 R^{n+1} 中不包含原点的非负真锥上是伪凸函数，其中矩阵

$$\begin{pmatrix} Q & a \\ a^T & 0 \end{pmatrix}$$

不包含全为 0 的行和列.

定理 3.17 设 Q 不包含全为 0 的行和列，则二次函数 $q(x)$ 在非负真锥 R_+^n 上是伪凸函数的等价条件是对于每一个 $x \in R_+^n$，存在函数 $r(x)$，$0 \leqslant r(x) < -\infty$，广义 Hesse 矩阵

$$H(x;r(x))=Q+r(x)(Qx+a)(Qx+a)^T$$

半正定.

定理 3.18 二次函数 $q(x)$ 在凸集 C 上是严格伪凸函数的等价条件是对于 $\{1,\cdots,n\}$ 的所有非空子集 P，$|D_P(x)| \leqslant 0$，$x \in C$. 而 $|D_P(x)|=0$ 时，Q 的顺序主子式均大为 0。

应该指出的是，不管是凸函数还是广义凸函数，它们的判别准则往往都需要检验无穷多个不等式的成立. 这在很多时候都是相当困难的. 所以发展其它类型的判别准则是广义凸分析的研究课题之一，目前已经有不少这方面的结果（见[4][10]）。

主要符号索引

R^n：n维欧氏空间.

$\|x\|$：向量x的模.

$d(x,y)$：向量x,y的距离.

$d(x,S)$：向量x与集合S的距离.

$\langle x,y\rangle$：向量x,y的内积.

$\varnothing$：空集.

$\dim S$：集合S的维数.

$\mathrm{aff}S$：集合S的仿射包.

$\mathrm{span}S$：集合S的线性包.

$\mathrm{co}S$：集合S的凸包.

$\overline{\mathrm{co}S}$：集合S的闭凸包.

$\overline{xy}$：连结向量x、y的线段.

$H(b,\alpha)$：超平面.

$\bar{K}(b,\alpha)$：以超平面$H(b,\alpha)$ 为边界的闭半空间.

$S^m(a_0,a_1,\cdots,a_m)$：以$a_0,a_1,\cdots,a_m$为顶点的单纯形.

$\mathrm{Lin}C$：凸集C的支撑子空间.

$\mathrm{int}S$：集合S的内部.

$\mathrm{ri}S$：集合S的相对内部.

$\mathrm{cl}S$：集合S的闭包.

$\mathrm{ext}C$：凸集C的极点集.

C°：凸集C的配极.

K^*：凸锥K的共轭锥.

$\tilde{K}$：凸锥K的极锥.

$\mathrm{cone}S$：集合S的锥包.

$B=\{x\,|\,d(x,\theta)\leqslant 1\}$：单位球.

A：R^n到R^m的线性变换.

A^*：线性变换A的伴随.

dom$f$：R^n中定义的函数$f(x)$ 的有效定义域.

epi$f$：R^n中定义的函数$f(x)$ 的上方图.

cl$f(x)$：R^n中定义的函数$f(x)$ 的闭包.

$\delta(x|C)$：凸集C的示性函数.

$\delta^*(x^*|C)$：凸集C的支撑函数.

$\gamma(x|C)$：凸集C的规范函数.

$f^*(x^*)$：函数$f(x)$ 的共轭函数.

$(f\lambda)(x)$，$\lambda\geqslant 0$：函数$f(x)$ 的非负右乘.

$(\lambda f)(x)$，$\lambda\geqslant 0$：函数 $f(x)$ 的非负左乘.

$(f_1+f_2)(x)$：函数$f_1(x)$，$f_2(x)$ 之和.

$(f_1\Box\cdots\Box f_m)(x)$：函数$f_1(x)$，$\cdots$，$f_m(x)$ 的下卷积.

co$g(x)$：函数$g(x)$ 的凸包.

co$\{f_i(x)|i\in I\}$：函数$f_i(x)$，$i\in I$的凸包.

$(Ah)(y)$：在线性变换A中，函数$h(x)$的象.

$(gA)(x)$：在线性变换A中，函数$g(y)$ 的逆象.

L-条件：Lipschitz 条件.

$\nabla f(x)$ 或$f'(x)$：函数$f(x)$ 的梯度.

$H(x)$ 或$f''(x)$：函数$f(x)$ 的Hesse 矩阵.

$f'(x;y)$：函数$f(x)$ 在x沿y方向的单边方向导数.

$\partial f(x)$：函数$f(x)$ 在x的次微分.

(LP)：线性规划.

(NLP)：非线性规划.

(CP)：凸规划.

(DNLP)：(NLP) 的Lagrange 对偶问题.

(SP)：鞍点问题.

$I(x)$：(NLP) 中，在x成为紧约束的不等式约束指标集.

$D(x)$：(NLP) 中，可行集X在x的可行方向锥.

$E(x)$：(NLP) 中，可行集X在x的线性化锥.

$F(x)$：目标函数$f(x)$ 在x的下降方向集合.

$G(x)$：$\{y | \langle y, \nabla g_i(x)\rangle < 0, i \in I(x)\}$.

$H_0(x)$：$\{y | \langle y, \nabla h_j(x)\rangle = 0, j = 1, \cdots, p\}$.

$T(x)$：可行集X在x的正切锥.

$G'(x)$：$\{y | \langle y, \nabla g_i(x)\rangle \leqslant 0, i \in I(x)\}$.

$A(x)$：可行集X在x的可达方向锥.

$L(x, \lambda, \mu)$：(CP) 的Lagrange 函数.

s.t.：*subjectto*，意为"约束条件是"。

参考书籍和文献

1. В. И. Аркин, В. Л. Лебин.
Выпуклость значений векторных интегралов, теоремы измеримого выбора и варнационные зауачи УМН 1972. 27.3. p 21-77.
2. J. P. Aubin
Applied Abstract Analysis John Wiley & Sons. Inc. New York 1977.
3. K. J. Arrow, A. C. Enthoven.
Quasi-Concave Programming. Economitrica 29.p779-800(1961).
4. M. Avriel
① *Nonlinear Programming. Analysis and methods. Prentice Hall. Englewood Cliffs. New Jersey.* 1976.
② *r-convex function. Math Prog* 2(1972) p309-323
5. M. Avriel, I. Zang
Generalized arcwize connected functions and characterizations of local-global minimum properties. J. Optimization Theory & Appl. 32 p407-425 (1980).
6. M. Avriel, S. Schaible.
Second order characterizations of pseudoconvex functions Math Prog 14 p170-185 (1978)
7. В. В. Береснев.
Об отображениях, сопряженных к выпуклым многозначныч отображениям. Кибернетика. 5(1973). p79-83.
8. V. Barbu, Th. Precupanu.
Convexity and Optimizations in Banach spaces. Editura Academici. România. 1978.
9. A. Ben-Tal.
On generalized means and generalized convex functions J. Optimization Theory & Appl 21(1977), p1-13.
10. C. Berge.
Topological Spaces. Oliver & Boyel Ltd Edinburgh. 1963.
11. C. Castaing.
Convex Analysis and Measurable Multifuctions. Lecture Notes in Mathematics. 580. *Springer-Verlag. Berlin.* 1977.

12. R. W. Cottle, J. A. Ferland.

① *Matrix-theoretic criteria for quasi-convexity and pseudo-convexity of quadratic functions. Linear Alg, Appl.* 5. p123-136 (1972).

② *On pseudo-convex functions of nonnegative variables. Math Prog.* 1 p95-101(1971).

13. J. P. Crouzeix.

① *Some differentiability properties of quasiconvex functions on R^n. Optimization and Optimal Control. Lecture Notes in Control and Information Sciences.* 30. *Springer-Verlag. Berlin.* 1981.

② *On second order conditions for quasiconvexity. Math Prog.* 18(1980) p349-352.

14. J. P. Crouzeix, J. A. Ferland.

Criteria for quasi-convexity and pseudoconvexity: relationship and comparisons. Math Prog. 23(1982). p193-205.

15. H. G. Eggleston.

Convexity Cambridge University Press. Cambridge. England. 1958.

16. А. Д. Иоффе, В, М. Тихомиров.

Теория экстремальных забач. М. Наука. 1974.

17. W. Fenchel.

Convex cones, sets and functions. Princeton University. Princeton. New Jersey 1951.

18. J. A. Ferland.

① *Matrix-theoretic criteria for the quasi-convexity of twice continuously differenciable functions. Linear Alg, Appl.* 38. p51-63(1981)

② *Mathematical programming problems with quasi-convex objective functions Math. Prog.* 3(1972) p296-301.

③ *Matrix criteria for pseudo-convex functions in the class C^2. Linear Alg, Appl.* 21. p47-57(1978).

19. W. Fleming.

Functions of Several Variables. Springer Verlag. New York. 1977.

20. H. J. Greenberg, W. P. Pierskalla.

① *Surrogate mathematical programs. OR.* 18. p924-939.(1970).

② *A review of quasi-convex functions. OR.* 19. p 1553-1570. (1971).

21. E. Hewitt, K. Stromberg.
Real and Abstract Analysis. Springer-Verlag. Berlin. 1978.

22. D. G. Luenberger.
Duality in mathematical programming. SIAMJ. Appl. Math. 16 p1090-1095 (1968).

23. O. L. Mangasarian.
① *Pseudo-convex functions J. SIAM. Control. Ser. A.* 3. p281-290(1965).
② *Nonlinear Programming. McGraw-Hill Book Co. New York* 1969.
③ *Convexity Pseudo-convexity and quasi-convexity of composite functions. Cahiers du centre d'Etude de Rech. Oper.* 12. p114-122 (1970).

24. B. Martos.
① *Subdefinite matrices and quadratic forms SIAMJ. Appl. Math.* 19. p1215-1233(1969).
② *Quadratic programming with quasi-convex objective function OR.* 19. p87-97(1971).

25. P. Mereau, J. G. Paquet.
Second order conditions for pseudo-convex functions. SIAM. J. Appl. Math. 27. p131-137(1974).

26. H. Nikaido.
On Von Neumann's minimax theorem. Pacific J. Math. 4. p65-72 (1954).

27. J Ponstein.
Seven kinds of convexity. SIAM Rew. 9. p115-119(1967).

28. Б. Н. Пшеничный
① *Выпуклый Анализ и Экстремальнне Задачи. М. Наука.* 1980.
② *Нообходимые условия экстремума. М. Наука* 1969.
③ *Выпуклые многозначные отображения и им сопряженные кибернетика* 1973. 3. p94-102.

29. R. T. Rockafellar.
① *Convex Analysis. Princeton University Press Princeton. New, Jersey* 1970.
② *Augmented lagrange multiplier functions and duality in nonconvex programming. SIAM. J. Control* 12.p268-285(1974)
③ *Level sets and continuity of conjugate convex functions Trans Amer Math. Soc.* 123. p46-63(1966).

30. S. Schaible.

① Q ty onvexity of cubic functions. *Math* (197.

② *S d order c ations of pseudo-convex quadratic funct ns. J. Optimization Theory & Appl.* 21. p15-26. (1977).

31. J. St, C. Witzgall.
Convexity and Optimization in Finite Dimensions. I. Springer-Verlag. Berlin 1970.
32. F. A. Valentine.
Convex sets. McGraw-Hill Book Co. New York. 1964.
33. 越民义.
凸分析讲义(上). 中国科学院应用数学研究所.1981.
34. 关肇直、张恭庆、冯德兴
线性泛函分析入门. 上海科技出版社 1979.
35. 王德人.
非线性方程组解法与最优化方法. 人民教育出版社. 1979.
36. 刘光中.
① 拟凸函数的一些性质. 成都科技大学学报 No.2. 1982.p.37~46.
② 广义凸规划中的*Lagrange*乘子及对偶. 成都科技大学学报. No.1 1983. p113-128.
③一类广义凸函数的判别准则.成都科技大学学报 *No.* 4. 1984. p19-24.
37. 王虹.
广义凸函数的一些判别准则. 成都科技大学应用数学系 78 级毕业论文集
38. 江泽涵.
拓扑学引论. 上海科学技术出版社. 1978.
39. И. П. Натансон.
Теория функции вещественной переменной. М. ИЛ. 1957.
40. M. S. Bazaraa, C. M. Shetty.
Nonlinear Programming. Theory and Algorithms. John Wiley & Sons. New York. 1979.
41. G. P. McCormick.
Second order conditions for constrained minima. SIAMJ. Appl. Math. 15. p641-652 (1967).

[General Information]

□□=□□□□□□□□

□□=□□□

□□=320

SS□=10069476

DX□=

□□□□=1991□05□□1□

□□□=□□□□□□□

3 [illegible]
4 [illegible]
[illegible]
[illegible] [illegible]
1 [illegible]
2 [illegible]
3 [illegible]
[illegible]
[illegible]
[illegible]